Universiteit Leiden

Opleiding Informatica

An Attempt to Detecting Patterns Among Children on the

Playground using Attributed Graph Mining

Name:	Alain Fonhof
Studentnr:	s1437690
Date:	22/08/2016
1st supervisor:	Matthijs van Leeuwen
2nd supervisor:	Ricardo Cachucho

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

An Attempt to Detecting Patterns Among Children on the Playground using Attributed Graph Mining

Alain Fonhof

Abstract

In this thesis we consider an attributed graph miner to find patterns among children who play on the playground. These patterns may provide us with a deeper understanding of the impact that the social-emotional skills of a child have on his social interactions. With a pattern mining approach we hope to find unexplored information that was not located by the previously used statistical approach. As our pattern mining method we chose CoPaM, an attributed graph miner that returns connected vertices with cohesive attributes. Firstly, we discuss the data pre-processing required to prepare the dataset as input for a dynamic social network whose vertices are associated with features. After that we examine the functionality and output of CoPaM. Next we visualize the output which gave an interesting insight into the interactions of children and provided a graphic overview of the data. Additionally, while analyzing the output of CoPaM we stumbled upon the fact that CoPaM was designed for a static instead of a dynamic attributed graph which caused a rise in the output of found patterns. To cope with this rise we focused on the frequency of each feature, the most prominent patterns and the pattern with the most vertices or features during post-processing of the output. In conclusion, a child's capacity to calm down or to be calmed down seemed to be the most prominent feature that was present in groups. Nevertheless, after visualizing and analyzing the output of CoPaM there seem to be no strong patterns in the data that present a correlation between social-emotional skills and social development.

Acknowledgements

I would first like to thank my supervisor Matthijs van Leeuwen for his valuable guidance. Additionally, I want to thank Guida Veiga and Ricardo Cachucho for their work and cooperation. Lastly, I thank Laurette Sonneville for correcting my English.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Pattern mining is a topic within data mining concerned with finding relevant patterns between data objects. Graph pattern mining methods expect a collection of graphs as input and return a set of subgraphs that satisfy some cohesion or density constraint. For our study the input graphs are undirected attributed graphs. This is relevant for social network analysis since a social network can be converted to an undirected attributed graph. An undirected edge between two vertices relates to a connection between two people in a social network. The vertices in the graph are labelled with features that give us information about a person in the network. A social network can be static or dynamic but in reality most social networks are not static since they change over time. [Berger-Wolf and Saia(2006)] describe how in a dynamic social network every timeframe can have a new list of edges or vertices.

In psychology one of the main research areas is about how people behave in a group environment. A group can be constructed as a social network if there is a relation between people in the group. In a behavioural paper such as [Veiga et al.(2016)Veiga, Ketelaar, Leng, Cachucho, Kok, Knobbe, Neto, and Rieffe] the authors research the impact of social-emotional skills on the social development of children ranging from the age of 4 to 6 years old. During these years the children develop social skills to interact and build up a relationship with their peers. One of the moments when children interact and play with each other is during recess time. When children prefer to spend their recess time alone instead of with their peers, then this is known to be an alarm signal for maladaptive social development as described by [Rubin et al.(2009)Rubin, Coplan, and Bowker] and [Coplan and Armer(2007)].

To research the impact of social-emotional skills on solitary behaviour, researchers observed 97 children on the playground during school recess time and tested their social-emotional skills. To observe these children, every child was given a Radio Frequency Identification Device (RFID) to measure their distance to other children. Every second the RFID-tag detected if a child was near another child and this was registered by the

RFID-tags. This registration is seen as an interaction between these children.

After collecting all the data, the researchers have processed the data with a statistical approach. This approach concluded that non-social behaviours do not necessarily evolve into later solitude. Nevertheless one non-social play that was associated with later solitude was solitary-pretend play (e.g., pretending to be a teacher alone), but only for girls and irrespective of their level of emotional competence. For our research we want to take a pattern mining approach on processing the data. The difference between a pattern mining approach and a statistical approach is that the statistical approach only shows us information about the interactions of children, whereas a pattern mining approach focuses more on the relationship between children. This gives us insight into how children interact and share similarities with each other. These patterns can be visualized in order to give a graphic overview of the interactions between children and give a deeper understanding of how they behave. In order to explore our approach we investigate and answer the following research question: Can we use attributed community detection to help psychologists understand the impact of social-emotional skills on the behaviour of children?

1.1 Approach

As our pattern mining algorithm we chose Cohesive Pattern Miner [Moser et al.(2009)Moser, Colak, Rafiey, and Ester]. CoPaM extracts cohesive patterns from an attributed graph. As input we are combining the children, their interactions and their features into an undirected social network whose vertices are linked with features. These features consist of the social-emotional skills and statistical analysis performed by the researchers. These two combined make up the feature space of the social network. Since the interactions of children cover a time period of 30 minutes we have to convert the social network to a dynamic social network. During pre-processing of the feature space we will have to make it compatible with the algorithm to find cohesive patterns. A cohesive pattern is a subgraph with similar attribute values. Figure 1.1 shows an example of a cohesive pattern where three children each have the value 1 for the social-emotional skill *Aggression*. This implies that these 3 children combined are a cohesive pattern and connected by their interactions and features. An attributed graph can contain multiple patterns, therefore during data post-processing we will study the most prominent patterns. With CoPaM we can find children who are connected through an interaction and their features. With this approach we hope to find a correlation between social-emotional skills and social interactions on a dynamic social network.

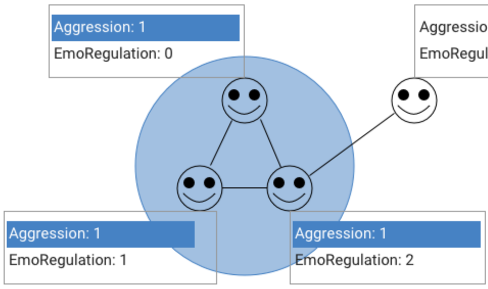

Figure 1.1: Example of a cohesive pattern

1.2 Main Contributions

- We analyzed the data about the social-emotional skills of children together with the domain expert Guida Veiga to improve data pre-processing.
- We shaped the feature space to be compatible with CoPaM by pre-processing the input data and discretizing the continuous variables.
- We implemented CoPaM based on the pseudo code in [Moser et al.(2009)Moser, Colak, Rafiey, and Ester].
- We constructed a dynamic attributed graph which served as input for CoPaM.
- We visualized the dynamic social network and the cohesive patterns that were the output of CoPaM.
- We analyzed the set of cohesive patterns that were extracted by CoPaM.

1.3 Thesis Overview

The rest of the paper is arranged as follows. Section 2 reviews related work. In Section 3, we discuss the input data for CoPaM. Section 4 explain the functionality of CoPaM and its output. Section 5 discusses the results of our experiment. Lastly, Section 6 concludes this paper with the findings and interesting topics for future research.

Chapter 2

Related Work

Social network analytics is in general context about interactions between people and determining the important structural patterns in such interactions. The authors [Aggarwal and Abdelzaher(2011)] discuss the key problems which arise with sensor data and the corresponding solutions. Also different dynamic models for a social network are suggested and analyzed. Extracting these structural patterns can be done in various ways and different techniques could result in different candidate graph patterns. [Akoglu et al.(2012)Akoglu, Tong, Meeder, and Faloutsos] present PICS (Parameter-free Identification of Cohesive Subgroups), a parameter-free clustering model to find groups of nodes with similar connectivity and attributed homogeneity. Since it is parameter-free it works without any user-specified input, such as the number of clusters, choice of density or similarity functions and thresholds. PICS constructs a matrix where users are carefully arranged to reveal the patterns. A pattern consists of nodes with a similar connectivity and high attribute homogeneity. In addition, it also clusters the attributes into attribute-clusters to show the distribution of attribute values. [Zhou et al.(2009)Zhou, Cheng, and Yu] produces a method that isolates a large graph whose vertices are associated with attributes into k clusters so that each cluster contains a densely connected subgraph with homogeneous attribute values called a SA-pattern. [Moser et al.(2009)Moser, Colak, Rafiey, and Ester] presents CoPaM (Cohesive Pattern Miner), which exploits various pruning strategies to efficiently find connected subgraphs whose vertices have cohesive attributes and filters patterns that are a subset of other found patterns.

In our application we are using the CoPaM algorithm. The reason behind this decision is that PICS works without parameters, which limits the filtering of false positive input data. With PICS a pattern has high attribute homogeneity, but for our problem we need the possibility to find patterns with specific attributes. SA-pattern was not chosen because a SA-pattern does not have to be a connected subgraph, though in our problem we are looking for groups of children that interact with each other. Therefore, a pattern has to be connected. Another reason why CoPaM is more relevant, is because it enables us to find common recurrent patterns.

Chapter 3

Data

In this chapter we will discuss the dataset that was used as input for our pattern miner.

3.1 Data acquisition

The vertices and edges of our social network consist of the children and interactions observed during the school recess time with the RFID-tags. The edges are undirected and have a lifetime of one second. The school recess time was 30 minutes on average and all interactions were stored in a dataset. This dataset is called the **interaction dataset** and consists of the children, adults and seesaws that have an interaction, their distance to each other and the timestamp of the interaction. Interactions with a seesaw is a special case since the distance between two children sitting on the ends of a seesaw can not be registered. Therefore, the ends of a seesaw were also given RFID-tags. If two children have an interaction with the ends of the same seesaw and have the same timestamp then this is counted as an interaction between these children. Table 3.1 shows the interaction dataset structure.

Column name	**Datatype**	**Description**
TagX	Numeric	Tag ID of child 1
TagY	Numeric	Tag ID of child 2
power	Numeric	Inverse distance between children
time	Timestamp	Timestamp of interaction

Table 3.1: Interaction dataset structure

[Veiga et al.(2016)Veiga, Ketelaar, Leng, Cachucho, Kok, Knobbe, Neto, and Rieffe] used the interaction dataset to extract information with a statistical approach. With their approach they derived additional data about behaviour of a child such as the number of interactions, mean time spent alone and percentage of time spent with same gender. This dataset is called the **derived interaction dataset** and its structure can be seen in Table 3.2

To determine the social-emotional skills of a child, children would participate in testing sessions. Besides testing sessions, the parents and teachers of the children filled out surveys. The results of these tests were collected in a dataset and is called the **social-emotional skills dataset**. The structure of the social-emotional skills dataset is shown in Table 3.3

Column name	**Datatype**	**Description**
child	Numeric	Tag ID
number.childs.interacted	Numeric	Number of different children interacted with
number.interactions	Numeric	Total number of interactions
mean.time.interactions	Numeric	Mean duration of an interaction
percentage.time.alone	Numeric	Percentage of their time alone
mean.count.simultaneous.interactions	Numeric	Mean count of simultaneous interactions
number.alone	Numeric	Number of time alone
mean.time.alone	Numeric	Mean time of their time alone
percentage.time.adult	Numeric	Percentage of their time with an adult
percentage.time.same.gender	Numeric	Percentage of their time with a child of the same gender
percentage.time.other.gender	Numeric	Percentage of their time with a child of the a different gender
percentage.time.interaction	Numeric	Percentage of their time making an interaction
percentage.time.interaction.one.child	Numeric	Percentage of their time interaction with only one child
percentage.time.interaction.group	Numeric	Percentage of their time interacting with multiple children
gender	String	Gender of the child
code	String	Child ID

Table 3.2: Derived interaction dataset structure

Column name	**Datatype**	**Description**
Code	String	Child ID
gender	Boolean	1 = male, 2 = female
age	Numeric	Age of parents
SocialCompTeachersMay	Numeric	Social Competence - rated by teachers in May (same time as Tags)
Empathy	Numeric	Empathy - rated by parents
EmoUnderstanding	Numeric	Emotion Understanding
Aggression	Numeric	Aggression
TheoryofMind	Numeric	Theory of Mind
SocialCompParentsDec	Numeric	Social Competence - rated by parents in December
SocialCompTeachersDec	Numeric	Social Competence - rated by teachers in December
PositiveEmotions	Numeric	Display of positive emotions
EmoRegulation	Numeric	Emotion Regulation
CapacitytoCalmDown	Numeric	Capacity to calm down or be calmed down
MotorCompetence	Numeric	Motor Competence
tagalonen	Numeric	Number of interactions alone
tagalonept	Numeric	Percentage of time alone
tagalonet	Numeric	Mean time alone
taggroupsize	Numeric	Group size - mean of simultaneous interactions
taginteractn	Numeric	Mean time of interactions

Table 3.3: Social-emotional skills dataset structuce

3.2 Data pre-processing

Generally real world data is either incomplete, noisy or inconsistent. The pattern mining algorithm looks for cohesive patterns in a connect subgraph that has homogeneous values in its feature space. Unfortunately, some features in the feature space have no homogeneous values. Therefore, in order to use our pattern mining algorithm we had to clean and discretize the gathered data. The feature space of our graph consists of the derived interaction dataset and social-emotional skills dataset.

3.2.1 Data cleaning

To clean the social-emotional skills dataset we first checked which children were present in the interaction datasets. Children that were not in any of the interaction datasets were removed from the social-emotional skills dataset. Secondly, we looked for children with missing values. Since some parents did not fill out the survey or children were absent at school during the social-emotional tests there were children with missing values. The following four children: BM280408, JP120208, MA090308 and MF110908 missed 6 data fields and were therefore excluded and removed from the social-emotional dataset. In addition, there were children who got registered with RFID-tags but did not have data about their social-emotional skills. Therefore the following two children: AP060507 and AS281207, were also excluded. Interactions with children who were excluded were ignored during graph initialisation.

Besides missing values there was also noisy data in the data collected by the RFID-tags. We checked the interaction datasets and rows with tag IDs that did not belong to any child, adult or seesaw were removed. In the `turma7afternoon` and `turma7morning` dataset the tags with ID 1329, 1652, 3927, 2174 and 4726 do not belong to any object. In the `turma8afternoon` dataset the tags with ID 808, 1056, 2744, 3911 and 43274 do not belong to any object.

3.2.2 Data reduction

In union, the interaction dataset and derived interaction dataset have a total of 30 features. The feature space that we use as input for our pattern mining algorithm should only consist of features that characterize the child. For that reason we want to reduce the number of features by removing irrelevant features. Firstly, the feature SocialCompTeachersMay missed 11 values and replacing it with the mean would make this feature irrelevant. Moreover, the feature Age represents the age of the parents of the child that filled in the survey and not the child itself. According to [Gottman and DeClaire(1997)] parenting skills yield the largest benefit to raising a child rather than age, so the feature age is irrelevant. Lastly, the features of the derived interaction dataset seemed to be valuable but during tests with the pattern mining algorithm we discovered that most patterns were found between children that share the same feature value of a feature from the derived interaction dataset. We discovered that the reason behind this is that these features were constructed based on the interaction dataset, which means that these features are based on the network structure and since the pattern mining algorithm is looking in the network for a connected subgraph with homogeneous values these features are redundant. The following features are removed from the feature space because of the aforementioned reasons:

- SocialCompTeachersMay
- Age
- number.childs.interacted
- number.interactions
- mean.time.interactions
- percentage.time.alone
- mean.count.simultaneous.interactions
- number.alone
- mean.time.alone
- percentage.time.adult
- percentage.time.same.gender
- percentage.time.other.gender
- percentage.time.interaction
- percentage.time.interaction.one.child
- percentage.time.interaction.group

3.2.3 Data discretization

Since the scale of measures used for the features is numerical there were some features with zero or close to zero homogeneous values. This level of measurement implies that the pattern mining algorithm can not use these features. In order to overcome this issue we discretized these features into nominal variables. To convert the features into nominal variables we have binned the feature values into four bins. We used the equal frequency binning technique on every feature except Gender, as this is a binary variable. We used the equal frequency technique because we want an equal distribution of children among every bin. This way there is no overpopulated bin which would result in the influx of patterns found. The range of every bin can be seen in Table 3.4. To indicate if feature value x should fall in bin number y the following equation should be true.

$$minrange(bin_y) < x \leq maxrange(bin_y) = T \tag{3.1}$$

feature	**0**	**1**	**2**	**3**
Empathy	[1.800, 2.700]	(2.700, 2.872]	(2.872, 3.337]	(3.337, 3.800]
EmoUnderstanding	[-2.481, -0.368]	(-0.368, 0.091]	(0.091, 0.631]	(0.631, 1.277]
Aggression	[1.000, 1.500]	(1.500, 1.917]	(1.917, 2.167]	(2.167, 4.000]
TheoryofMind	[1.000, 1.4999]	(1.4999, 1.500]	(1.500, 1.667]	(1.667, 2.000]
SocialCompParentsDec	[2.000, 2.429]	(2.429, 2.690]	(2.690, 2.857]	(2.857, 3.000]
SocialCompTeachersDec	[1.571, 2.286]	(2.286, 2.429]	(2.429, 2.714]	(2.714, 3.000]
PositiveEmotions	[3.000, 3.833]	(3.833, 4.333]	(4.333, 4.792]	(4.792, 5.000]
EmoRegulation	[1.125, 2.000]	(2.000, 2.375]	(2.375, 2.625]	(2.625, 3.500]
CapacitytoCalmDown	[2.333, 3.333]	(3.333, 3.667]	(3.667, 4.000]	(4.000, 5.000]
MotorCompetence	[-1.536, -0.500]	(-0.500, 0.072]	(0.072, 0.437]	(0.437, 1.265]

Table 3.4: Bins sizes of the numerical variables

Chapter 4

Methods

In this chapter we will discuss the CoPaM algorithm and how we want to apply this pattern miner to our network.

4.1 CoPaM

With our application of CoPaM we are researching the groups of children and what features these groups share. Besides the connections between children, which a cluster algorithm such as [Zhang(2005)] researches, the patterns will provide us with information about features that correlate with interactions. That is, if we find frequent patterns with a the same feature and feature value this could imply that this is an important feature for social development. The implementation of CoPaM was based on the pseudo code provided by [Moser et al.(2009)Moser, Colak, Rafiey, and Ester]. CoPaM is a pattern mining algorithm that exploits various pruning strategies to find all maximal cohesive patterns in a network. The input of CoPaM is a feature vector graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{D})$. A feature vector graph is a graph that consists of vectors that each have a feature value for every feature in the feature space $\mathcal{D}$. For the feature vector graph definition see Definition 1 by [Moser et al.(2009)Moser, Colak, Rafiey, and Ester]). A cohesive pattern is an induced subgraph $G = (V, E, D), V \subset \mathcal{V}, E = \{v_1, v_2 | v_1, v_2 \in V, \{v_1, v_2\} \in \mathcal{E}\}, D \subset \mathcal{D}$ that satisfies the following three constraints:

- **Subspace cohesion constraint:** *G is homogeneous in* $D \subseteq \mathcal{D}$, *i.e.* $s(V, D, \theta_s) = true$ *and* $|D| \geq \theta_{\text{dim}} \geq 1$
- **Density constraint:** $d(G) := \frac{2|E|}{|V|(|V|-1)} \geq \alpha$. *(In this case G is also called* α*-dense.)*
- **Connectivity constraint:** *G is connected.*

These three constraints combined are called the cohesive pattern constraint (**CP constraint**). In addition, an edge is called cohesive if the induced graph, constructed with its two connected nodes, fulfills the CP

Figure 4.1: Example of a maximal cohesive pattern

constraint. Figure 1.1 shows a cohesive pattern that complies with the CP constraint.

The output of the CoPaM algorithm is the set of maximal cohesive patterns. A maximal cohesive pattern can be defined as a cohesive pattern that can not be extended by a node while still being a cohesive pattern. Furthermore, the feature subspace of a cohesive pattern should be maximal, which implies that there does not exist an additional feature whose feature values are homogeneous among the nodes of the pattern. For the maximal cohesive pattern definition see Definition 3 by [Moser et al.(2009)Moser, Colak, Rafiey, and Ester]. Figures 4.1 shows a toy example of a maximal cohesive pattern. Both patterns, left and right, are a cohesive pattern but the left pattern is not a maximal cohesive pattern because it can be extended by a vertex and still comply with the CP constraint.

4.1.1 Feature vector graph

The feature vector graph $\mathcal{G}$ is constructed with a node set $\mathcal{V}$, edge set $\mathcal{E}$ and feature space $\mathcal{D}$. We will discuss how we implemented each part. The feature space $\mathcal{D}$ is discussed in Chapter 3 and consists of the following eleven features:

- gender
- Empathy
- EmoUnderstanding
- Aggression
- TheoryofMind
- SocialCompParentsDec
- SocialCompTeachersDec
- PositiveEmotions
- EmoRegulation
- CapacitytoCalmDown
- MotorCompetence

The node set $\mathcal{V}$ consists of the children that were present in the interaction dataset minus the children whose data were incomplete (see 3.2.1). The edge set $\mathcal{E}$ consists of the registered interactions between children in the interaction dataset. The interaction dataset registered 30 minutes of interactions, which causes multiple edges between two vertices. In order to maintain an overview of the data we introduced a dynamic social network.

DEFINITION 1. *[Dynamic social network] Let $\mathcal{N} = (T, \theta_i, \mathcal{G})$ be a dynamic social network with the following parameters:*

- *Time interval in seconds T,*
- Interaction threshold θ_i,
- *Feature vector graph $\mathcal{G}$*

We study our dynamic social network across the time axis and assume a series of discrete timepoints $t_1, \ldots, t_{n-1}, t_n$. For our application we only change the edge set $\mathcal{E}$ over time, so at timepoint t_i we observe the feature vector graph instance $\mathcal{G}_i = (\mathcal{V}, \mathcal{E}_i, \mathcal{D})$. The interval T is the number of seconds between every timeframe. Every T seconds the edge set is cleared and filled with the interactions of the following T seconds. For our application we chose $T = 60$ because increasing T reduces the change of disconnected vertices and decreasing T would increase the number of disconnected vertices. An alternative would be a sliding window to show a graph representation of every second. We did not implement a sliding window because it would enlarge output of maximal cohesive patterns and this would clutter the results. To exclude false positive interactions, for example children that briefly brush up against each other, we set an interaction threshold. The interaction threshold θ_i implies that there should be a minimum of θ_i edges between two nodes during an interval. Thus if two children have θ_i or more interactions its edge is included in the edge set $\mathcal{E}$. For our application we set $\theta_i = 5$ so that two children need to have $\frac{1}{12} \cdot T$ or more interactions during an interval. Since two children can have more than one interaction during an interval we assign a weight to every edge. The weight w_E of edge E relates to the number of interactions between two children.

4.1.2 CoPaM parameters

CoPaM accepts the following parameters:

- *density threshold α, $\frac{1}{3} < \alpha \leq 1$,*
- *subspace cohesion function s,*
- *subspace cohesion threshold θ_s,*
- *dimensionality threshold $\theta_d im$*

For our implementation we are using the following parameters. Density threshold $\alpha = \frac{1}{2}$ since $\alpha = 1$ would imply that every child in a pattern is a maximum of 1,5 meter away from each other. Density threshold $\alpha = \frac{2}{3}$ and $\alpha = \frac{1}{2}$ increase this maximum distance between two children in a pattern to $(|V| - 1) \cdot 1,5m$. Though both are valid options we do not want to exclude loosely organized groups of playing children, so we set our density threshold to $\alpha = \frac{1}{2}$. We are not using density threshold $\alpha = \frac{1}{3}$ since our node set consists of approximately 20 nodes and a density threshold of $\alpha = \frac{1}{3}$ would need a minimum of 6 vectors in a pattern to be a connected subgraph making this irrelevant for our application. We chose the following subspace cohesion function s:

$$s(V, D, true) = \forall v \in V, d \in D : F_d(v) = true \tag{4.1}$$

$F_d(v)$ stands for the feature value of node v in dimension d. The subspace cohesion function requires that every node in a pattern has a identical feature value in dimension D. For example, in Figure 1.1 three nodes share the same feature value for the feature Aggression. Since we are not using the subspace cohesion threshold parameter this parameter can be set to $\theta_s = true$. As input for our minimum dimensions parameter we chose $\theta_{dim} = 3$ which implies that dimension $|D| \geq 3$. Increasing the minimum number of dimensions threshold causes overfitting considering only a few children share 5 or more features values. Table 4.1 shows the average maximal cohesive patterns found for each dataset. The average is calculated by dividing the sum of the maximal cohesive patterns per timeframe t by the total number of timeframes.

$$averagemaximalcohesive = \frac{\sum_{i=0}^{n} |maximal\ cohesive\ patterns_i|}{n} \tag{4.2}$$

Minimum dimensions $\theta_{dim} \geq 5$ results in some timeframes having no cohesive patterns and is therefore not viable.

θdim	turma6-afternoon	turma6-morning	turma7-afternoon	turma7-morning	turma8-afternoon	turma8-morning	turma9-afternoon	turma9-morning
2	23	26	17	17	10	13	4	5
3	19	23	15	16	10	15	5	6
4	13	16	11	11	8	12	4	6

Table 4.1: Average maximal cohesive patterns per interaction dataset

Chapter 5

Experiments

We evaluate CoPaM on being able to present patterns that display a positive or negative correlation between features and interactions. We visualize and analyze the generated output of the input described in Chapter 4. The input data for our experiment reflects children playing on the playground. During our experiment we are looking to identify groups of children that play together. When analyzing these groups we hope to find features and feature values that are frequently present since this would imply a correlation between these features and social interaction. Our dynamic networks have an average of 30 timeframes. Figure 5.1 shows timeframe t_3 from social network `turma6afternoon` with the node color decided by the node feature 'gender'. When the gender equals male, female or adult the color is either blue, pink or red respectively. For the full visualization of the social networks see [Fonhof(2016)]. In this visualization we highlighted 3 patterns that were acquired during postprocessing. Figure 5.2 shows a snapshot from the visualization. To clarify the visualization and the 3 patterns:

- In the lop left corner we see a normal network without any highlight. This image shows how the different genders interact with each other.
- In the top right corner we see the network with a green highlighted pattern. This pattern is at that time the maximal cohesive pattern with the most children together, $\max |V|$.
- In the bottom left corner we see the network with a yellow highlighted pattern. This pattern is at that time the maximal cohesive pattern where the children share the most number of features, $\max |D|$.
- In the bottom right corner we see the network with an orange highlighted pattern. This pattern is at that time the maximal cohesive pattern pattern with the highest value for the equation: $|D| \cdot |V|$.

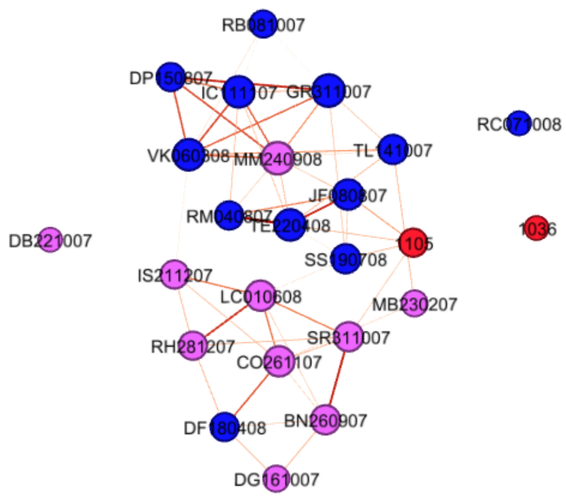

Figure 5.1: Example of a network timeframe

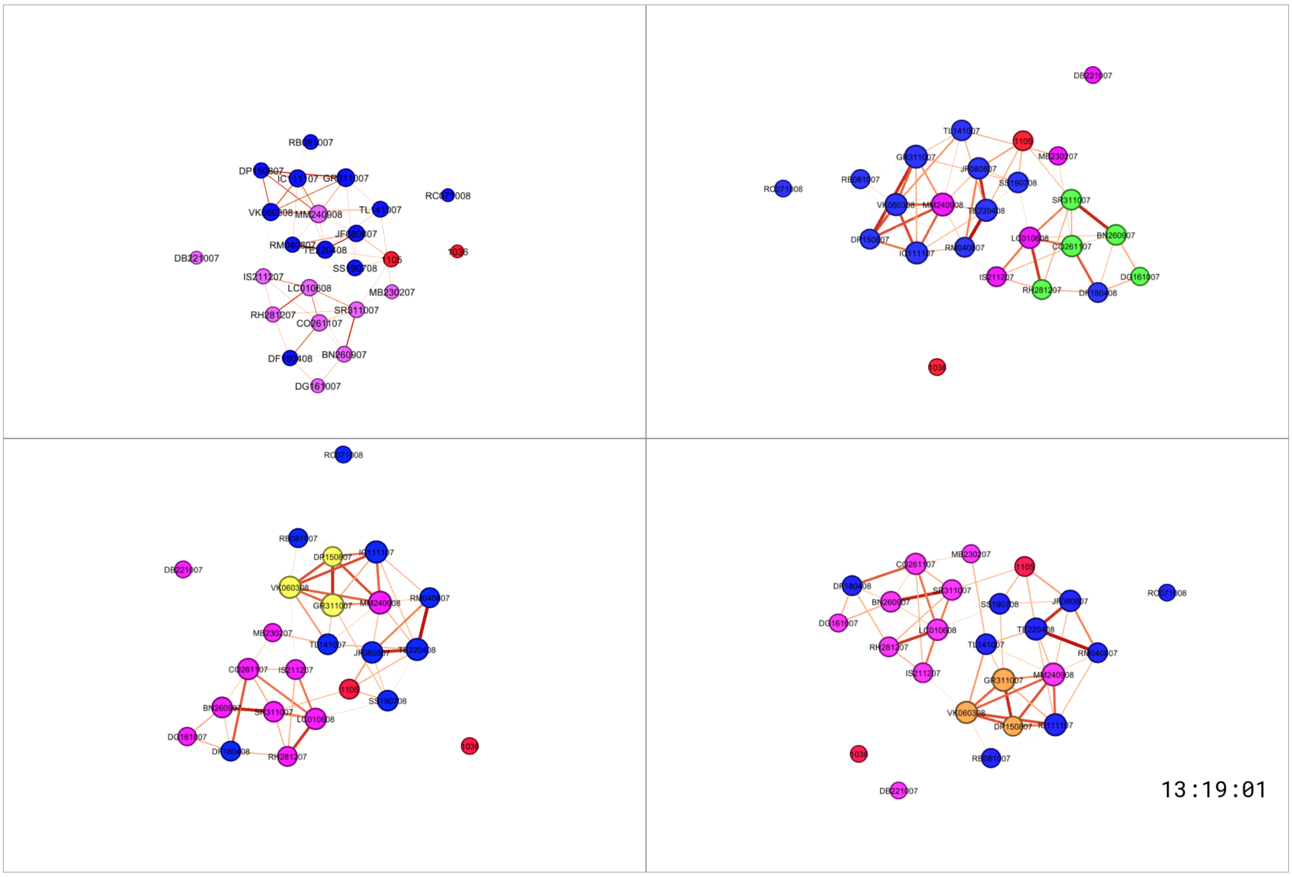

Figure 5.2: Snapshot of the visualization

5.1 Results

Table A.1 shows the number of maximal cohesive patterns for every snapshot per dynamic network. This overview displays the increase of patterns found by converting the social network to a dynamic social network. Given that we found over 100 maximal cohesive patterns for every single dynamic network we decided to first provide further inspection of the composition of the found patterns.

Table A.2 presents the most frequent maximal cohesive pattern for each dynamic network. From the 8 presented patterns, 7 are unique combinations of children. Each pattern seems to be friends of the same gender playing together, though `turma7afternoon` dataset shows a boy and a girl interacting. While analyzing the boy GQ110608 and girl NL201207 in the visualization we can see them either playing with each other or a group of girls.

Table A.3-A.9 shows the frequency of each feature and its respective bin for each dynamic network. An explanation for why gender 1(male) has more patterns could be that the distribution is skewed towards boys. There are a total of 44 boys to 26 girls in the social-emotional skills dataset. TheoryofMind with bin [1.49, 1.5] is a frequent combination, although this could also be incited by the distribution of this feature. Considering there are 28 children with the value 1,5 for TheoryofMind, an even distribution is not possible.

Table A.11-A.18 shows the frequency of each feature per dynamic network. These tables reveal that gender is the most important feature, however this is caused by the fact that gender is a binary value and thus only has 2 bins compared to the 4 bins of the other features. When we review the ranking of each feature across all datasets (see Table A.19) there are a few noteworthy features such as CapacitytoCalmDown and MotorCompetence. CapacitytoCalmDown because its lowest frequency rank across all datasets is 4th and MotorCompetence for its highest frequency rank being 7th. Which hints that children with the same value for CapacitytoCalmDown group up with each other and also children with equal values for MotorCompetence do not seem to strive for interactions.

Chapter 6

Conclusions and further research

In this paper we examined the application of the pattern mining algorithm CoPaM to detect features that impact the social interactions of a child on the playground. While analyzing the output of CoPaM the first complication was the fact that we were working with a dynamic social network. There was a linear relationship between the amount of patterns found and the number of snapshots taken which caused an increase of total patterns found per network. This clutters the output and raises the importance of postprocessing to filter the irrelevant patterns. By analyzing the output we could not determine a clear correlation or noncorrelation between a feature and the social interactions of a child. However, CapacitytoCalmDown seemed to consistently be one of the top 4 features and MotorCompetence one of the bottom 3 features. This could imply that children with the same value for CapacitytoCalmDown tend to group up together and that MotorCompetence has little impact. The psychologists were also interested in children who show signs of maladaptive social development by spending their time alone instead of with their peers, but CoPaM can not detect these patterns because a disconnected vertex can not be a maximal cohesive pattern and is therefore not included in the output. To conclude, CoPaM does not seem to be a suitable algorithm for mining patterns on a dynamic feature vector graph and did not find any strong patterns that present a correlation between social-emotional skills and social development. The biggest problem is the lack of dynamic analysis, a possible solution to this problem is evolutionary clustering over time by [Chakrabarti et al.(2006)Chakrabarti, Kumar, and Tomkins] or ContextTour [Lin et al.(2010)Lin, Sun, Cao, and Liu] which consists of the two components: Dynamic Relational Clustering and Dynamic Network Countour-map.

In this paper we have mainly focused on patterns that share the same features. To further study social-emotional skills in a dynamic social network one could research the patterns of not only connected but also disconnected vertices and analyze the difference in features and behaviour. Another point that could be explored is disconnected patterns. The output of CoPaM are connected patterns but a pattern of vertices with homogeneous features and how they behave could be interesting. Furthermore solutions to the problems that arise with the upsurge of patterns while pattern mining a dynamic social network would be fruitful.

Appendix A

Output analysis

Timeframe t_i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Turma6afternoon	40	36	20	26	24	30	20	23	19	20	21	12	11	15	21	22	15	25	19	19	15	13	17	15	20	21	16	10	11	17	-
Turma6morning	26	24	23	23	19	16	21	22	16	29	30	23	23	23	21	20	22	19	24	26	23	25	24	20	28	30	24	21	28	19	-
Turma7afternoon	20	15	16	17	19	14	17	18	16	14	22	19	15	16	16	14	14	18	15	18	20	17	8	15	15	16	11	8	11	10	-
Turma7morning	21	19	15	16	19	20	17	20	21	23	22	19	12	14	15	16	9	13	15	15	11	13	12	13	13	18	17	16	16	18	-
Turma8afternoon	16	9	9	11	13	10	11	13	11	13	6	9	13	13	11	16	11	7	14	7	8	9	10	12	10	12	7	10	6	7	-
Turma8morning	18	21	16	17	17	20	14	16	17	16	18	11	14	16	12	15	12	9	9	13	15	11	16	21	19	16	18	21	13	14	14
Turma9afternoon	1	3	3	4	4	4	5	4	6	7	8	7	4	4	3	3	5	6	6	7	6	10	5	7	6	5	4	7	4	5	-
Turma9morning	7	4	4	8	10	8	9	7	7	7	7	7	8	5	5	5	3	5	8	9	5	5	6	7	6	8	4	4	-	-	-

Table A.1: Maximal Cohesive Patterns per timeframe for every interaction dataset

	turma6-afternoon	turma6-morning	turma7-afternoon	turma7-morning	turma8-afternoon	turma8-morning	turma9-afternoon	turma9-morning
Pattern Children	DP150807, GR311007, VK060308	TL141007, VK060308	GQ110608, NL201207	AT200408, MF301207	AS070208, EC0505082	LF100408, MR121008	BO190708, CE051208	BO190708, CE051208
Number of Children	3	2	2	2	2	2	2	2
Pattern frequency	23	27	21	19	25	26	23	22
Number of cohesive features	3	4	3	5	5	4	4	4
Cohesive features	'TheoryofMind': '(1.49, 1.5]', 'gender': 1, 'Aggression': '(2.167, 4]'	'TheoryofMind': '(1.49, 1.5]', 'gender': 1, 'MotorCompetence': '(0.437, 1.265]', 'CapacitytoCalmDown': '(3.333, 3.667]'	'Empathy': '(2.872, 3.337]', 'MotorCompetence': '(-0.5, 0.072]', 'PositiveEmotions': '(3.833, 4.333]'	'TheoryofMind': '(1.49, 1.5]', 'gender': 1, 'SocialCompTeachersDec': '[1.571, 2.286]', 'EmoRegulation': '(2.375, 2.625]', 'SocialCompParentsDec': '(2.69, 2.857]'	'EmoRegulation': '(2, 2.375]', 'gender': 1, 'CapacitytoCalmDown': '(3.667, 4]', 'Aggression': '(1.5, 1.917]', 'SocialCompParentsDec': '(2.69, 2.857]'	'TheoryofMind': '(1, 1.49]', 'gender': 2, 'Aggression': '(1.5, 1.917]', 'CapacitytoCalmDown': '(3.667, 4]'	'CapacitytoCalmDown': '(3.333, 3.667]', 'gender': 2, 'SocialCompTeachersDec': '[1.571, 2.286]', 'SocialCompParentsDec': '(2.429, 2.69]'	'CapacitytoCalmDown': '(3.333, 3.667]', 'gender': 2, 'SocialCompTeachersDec': '[1.571, 2.286]', 'SocialCompParentsDec': '(2.429, 2.69]'''

Table A.2: Most frequent pattern per interaction dataset

feature	bin	frequency
gender	1	296
TheoryofMind	(1.49, 1.5]	141
SocialCompTeachersDec	(2.714, 3]	119
CapacitytoCalmDown	(3.667, 4]	102
gender	2	100
SocialCompParentsDec	(2.69, 2.857]	80
EmoUnderstanding	[-2.481, -0.368]	78
SocialCompTeachersDec	[1.571, 2.286]	68
Empathy	(3.337, 3.8]	66
PositiveEmotions	(3.833, 4.333]	61
TheoryofMind	(1.5, 1.67]	61
Empathy	[1.8, 2.7]	56
EmoUnderstanding	(-0.368, 0.0914]	54
MotorCompetence	(0.072, 0.437]	53
Aggression	(2.167, 4]	50
EmoRegulation	(2, 2.375]	48
SocialCompParentsDec	[2, 2.429]	44
MotorCompetence	(-0.5, 0.072]	43
EmoRegulation	(2.375, 2.625]	41
Aggression	[1, 1.5]	37
MotorCompetence	[-1.536, -0.5]	33
EmoUnderstanding	(0.631, 1.277]	31
CapacitytoCalmDown	(3.333, 3.667]	30
PositiveEmotions	[3, 3.833]	28
Aggression	(1.917, 2.167]	28
EmoRegulation	[1.125, 2]	28
SocialCompParentsDec	(2.429, 2.69]	27
PositiveEmotions	(4.333, 4.792]	25
SocialCompTeachersDec	(2.429, 2.714]	24
CapacitytoCalmDown	[2.333, 3.333]	24
CapacitytoCalmDown	(4, 5]	22
Empathy	(2.7, 2.872]	22
SocialCompParentsDec	(2.857, 3]	21
PositiveEmotions	(4.792, 5]	21
MotorCompetence	(0.437, 1.265]	20
EmoRegulation	(2.625, 3.5]	17
Empathy	(2.872, 3.337]	15
TheoryofMind	(1.67, 2]	9
EmoUnderstanding	(0.0914, 0.631]	8
Aggression	(1.5, 1.917]	7
TheoryofMind	(1.0, 1.49]	3

Table A.3: The frequency of each feature bin for turma6afternoon

feature	bin	frequency
gender	1	373
TheoryofMind	(1.49, 1.5]	195
SocialCompTeachersDec	(2.714, 3]	132
gender	2	116
CapacitytoCalmDown	(3.667, 4]	92
SocialCompParentsDec	(2.69, 2.857]	79
SocialCompTeachersDec	[1.571, 2.286]	79
TheoryofMind	(1.5, 1.67]	73
EmoRegulation	[1.125, 2]	70
EmoUnderstanding	[-2.481, -0.368]	69
Aggression	[1, 1.5]	63
CapacitytoCalmDown	(3.333, 3.667]	63
Aggression	(1.917, 2.167]	63
Empathy	[1.8, 2.7]	60
PositiveEmotions	(3.833, 4.333]	56
EmoUnderstanding	(-0.368, 0.0914]	55
Empathy	(3.337, 3.8]	53
EmoRegulation	(2.375, 2.625]	46
MotorCompetence	(-0.5, 0.072]	44
Empathy	(2.7, 2.872]	44
PositiveEmotions	[3, 3.833]	40
MotorCompetence	(0.437, 1.265]	39
EmoUnderstanding	(0.631, 1.277]	38
SocialCompParentsDec	[2, 2.429]	35
MotorCompetence	(0.072, 0.437]	35
MotorCompetence	[-1.536, -0.5]	34
SocialCompParentsDec	(2.429, 2.69]	33
Aggression	(2.167, 4]	32
TheoryofMind	(1.67, 2]	28
EmoRegulation	(2, 2.375]	28
EmoRegulation	(2.625, 3.5]	24
PositiveEmotions	(4.792, 5]	24
CapacitytoCalmDown	[2.333, 3.333]	24
SocialCompParentsDec	(2.857, 3]	23
Aggression	(1.5, 1.917]	22
SocialCompTeachersDec	(2.429, 2.714]	21
Empathy	(2.872, 3.337]	13
CapacitytoCalmDown	(4, 5]	12
EmoUnderstanding	(0.0914, 0.631]	12
PositiveEmotions	(4.333, 4.792]	11
TheoryofMind	(1.0, 1.49]	2

Table A.4: The frequency of each feature bin for turma6morning

feature	**bin**	**frequency**
gender	1	214
CapacitytoCalmDown	[2.333, 3.333]	139
TheoryofMind	(1.49, 1.5]	83
SocialCompParentsDec	[2, 2.429]	78
SocialCompTeachersDec	(2.429, 2.714]	76
EmoRegulation	(2.625, 3.5]	75
gender	2	73
Aggression	[1, 1.5]	71
Empathy	[1.8, 2.7]	61
Empathy	(2.872, 3.337]	58
MotorCompetence	(-0.5, 0.072]	54
EmoUnderstanding	(0.631, 1.277]	53
Aggression	(2.167, 4]	53
PositiveEmotions	(4.792, 5]	49
Empathy	(3.337, 3.8]	45
MotorCompetence	[-1.536, -0.5]	35
PositiveEmotions	(3.833, 4.333]	35
SocialCompTeachersDec	[1.571, 2.286]	34
PositiveEmotions	[3, 3.833]	32
EmoRegulation	(2, 2.375]	30
EmoUnderstanding	[-2.481, -0.368]	29
Empathy	(2.7, 2.872]	28
MotorCompetence	(0.072, 0.437]	27
CapacitytoCalmDown	(3.333, 3.667]	26
TheoryofMind	(1, 1.49]	26
EmoRegulation	(2.375, 2.625]	25
SocialCompParentsDec	(2.857, 3]	21
EmoUnderstanding	(0.0914, 0.631]	20
TheoryofMind	(1.5, 1.67]	20
SocialCompTeachersDec	(2.286, 2.429]	19
EmoUnderstanding	(-0.368, 0.0914]	18
PositiveEmotions	(4.333, 4.792]	18
SocialCompParentsDec	(2.69, 2.857]	16
Aggression	(1.917, 2.167]	15
CapacitytoCalmDown	(4, 5]	11
EmoRegulation	[1.125, 2]	11
MotorCompetence	(0.437, 1.265]	7
TheoryofMind	(1.67, 2]	5

Table A.5: The frequency of each feature bin for turma7afternoon

feature	**bin**	**frequency**
gender	1	265
CapacitytoCalmDown	[2.333, 3.333]	144
TheoryofMind	(1.49, 1.5]	98
gender	2	90
SocialCompTeachersDec	(2.429, 2.714]	87
Empathy	[1.8, 2.7]	82
PositiveEmotions	(4.792, 5]	74
Aggression	[1, 1.5]	69
EmoRegulation	(2.625, 3.5]	67
Empathy	(3.337, 3.8]	66
SocialCompParentsDec	[2, 2.429]	62
EmoRegulation	(2, 2.375]	52
SocialCompTeachersDec	[1.571, 2.286]	51
MotorCompetence	(-0.5, 0.072]	50
Aggression	(2.167, 4]	50
SocialCompTeachersDec	(2.286, 2.429]	44
EmoUnderstanding	(0.0914, 0.631]	44
PositiveEmotions	[3, 3.833]	37
SocialCompParentsDec	(2.69, 2.857]	37
EmoUnderstanding	(0.631, 1.277]	36
MotorCompetence	[-1.536, -0.5]	32
PositiveEmotions	(4.333, 4.792]	32
EmoRegulation	(2.375, 2.625]	27
CapacitytoCalmDown	(3.333, 3.667]	24
MotorCompetence	(0.072, 0.437]	24
Empathy	(2.872, 3.337]	21
SocialCompParentsDec	(2.857, 3]	20
TheoryofMind	(1.67, 2]	20
TheoryofMind	(1, 1.49]	20
EmoUnderstanding	(-0.368, 0.0914]	19
EmoUnderstanding	[-2.481, -0.368]	18
MotorCompetence	(0.437, 1.265]	17
CapacitytoCalmDown	(4, 5]	14
TheoryofMind	(1.5, 1.67]	14
Empathy	(2.7, 2.872]	12
Aggression	(1.917, 2.167]	11
PositiveEmotions	(3.833, 4.333]	9
EmoRegulation	[1.125, 2]	6
Aggression	(1.5, 1.917]	1

Table A.6: The frequency of each feature bin for turma7morning

feature	bin	frequency
gender	1	256
CapacitytoCalmDown	(3.667, 4]	87
EmoRegulation	(2, 2.375]	86
TheoryofMind	(1.49, 1.5]	59
Aggression	(1.5, 1.917]	55
MotorCompetence	(0.437, 1.265]	38
Empathy	(3.337, 3.8]	36
PositiveEmotions	[3, 3.833]	35
EmoRegulation	(2.625, 3.5]	34
SocialCompParentsDec	(2.69, 2.857]	33
SocialCompParentsDec	[2, 2.429]	33
SocialCompTeachersDec	[1.571, 2.286]	31
Aggression	(1.917, 2.167]	28
gender	2	27
TheoryofMind	(1, 1.49]	27
Empathy	(2.7, 2.872]	27
EmoRegulation	[1.125, 2]	27
TheoryofMind	(1.67, 2]	25
EmoUnderstanding	(-0.368, 0.0914]	24
EmoUnderstanding	[-2.481, -0.368]	23
EmoUnderstanding	(0.631, 1.277]	20
MotorCompetence	(0.072, 0.437]	18
SocialCompParentsDec	(2.429, 2.69]	16
PositiveEmotions	(3.833, 4.333]	15
CapacitytoCalmDown	[2.333, 3.333]	13
CapacitytoCalmDown	(3.333, 3.667]	10
SocialCompTeachersDec	(2.714, 3]	10
PositiveEmotions	(4.333, 4.792]	10
EmoUnderstanding	(0.0914, 0.631]	8
SocialCompTeachersDec	(2.429, 2.714]	7
Aggression	(2.167, 4]	7
MotorCompetence	[-1.536, -0.5]	6
Empathy	[1.8, 2.7]	5
Aggression	[1, 1.5]	4
Empathy	(2.872, 3.337]	4
CapacitytoCalmDown	(4, 5]	3
SocialCompTeachersDec	(2.286, 2.429]	3

Table A.7: The frequency of each feature bin for turma8afternoon

feature	bin	frequency
gender	1	340
TheoryofMind	(1.49, 1.5]	116
CapacitytoCalmDown	(3.667, 4]	113
EmoRegulation	(2, 2.375]	102
SocialCompTeachersDec	[1.571, 2.286]	83
Aggression	(1.5, 1.917]	75
Empathy	(3.337, 3.8]	66
gender	2	62
TheoryofMind	(1, 1.49]	62
SocialCompParentsDec	(2.69, 2.857]	60
PositiveEmotions	[3, 3.833]	58
Aggression	(1.917, 2.167]	58
EmoRegulation	(2.625, 3.5]	46
MotorCompetence	(0.437, 1.265]	43
Empathy	(2.7, 2.872]	43
SocialCompTeachersDec	(2.714, 3]	34
EmoUnderstanding	(-0.368, 0.0914]	34
MotorCompetence	[-1.536, -0.5]	33
EmoUnderstanding	[-2.481, -0.368]	29
Empathy	(2.872, 3.337]	28
SocialCompParentsDec	[2, 2.429]	25
CapacitytoCalmDown	[2.333, 3.333]	24
TheoryofMind	(1.67, 2]	20
SocialCompParentsDec	(2.429, 2.69]	20
SocialCompTeachersDec	(2.429, 2.714]	19
EmoUnderstanding	(0.0914, 0.631]	19
PositiveEmotions	(3.833, 4.333]	18
MotorCompetence	(0.072, 0.437]	15
Empathy	[1.8, 2.7]	13
EmoUnderstanding	(0.631, 1.277]	12
EmoRegulation	[1.125, 2]	12
Aggression	[1, 1.5]	11
SocialCompTeachersDec	(2.286, 2.429]	9
CapacitytoCalmDown	(3.333, 3.667]	6
PositiveEmotions	(4.333, 4.792]	6
CapacitytoCalmDown	(4, 5]	3
Aggression	(2.167, 4]	1

Table A.8: The frequency of each feature bin for turma8morning

feature	bin	frequency
SocialCompTeachersDec	[1.571, 2.286]	114
gender	1	58
gender	2	53
EmoRegulation	[1.125, 2]	47
EmoUnderstanding	(-0.368, 0.0914]	37
CapacitytoCalmDown	(4, 5]	36
CapacitytoCalmDown	(3.333, 3.667]	29
SocialCompParentsDec	(2.857, 3]	27
SocialCompParentsDec	(2.429, 2.69]	23
SocialCompTeachersDec	(2.429, 2.714]	22
EmoRegulation	(2, 2.375]	18
PositiveEmotions	(4.792, 5]	16
Empathy	(2.872, 3.337]	16
SocialCompParentsDec	(2.69, 2.857]	15
Empathy	[1.8, 2.7]	15
TheoryofMind	(1.67, 2]	9
MotorCompetence	[-1.536, -0.5]	8
MotorCompetence	(0.437, 1.265]	8
Aggression	[1, 1.5]	6
MotorCompetence	(-0.5, 0.072]	6
Aggression	(1.5, 1.917]	6
PositiveEmotions	(3.833, 4.333]	6
EmoUnderstanding	[-2.481, -0.368]	5
Aggression	(1.917, 2.167]	5
Aggression	(2.167, 4]	4
TheoryofMind	(1.49, 1.5]	3
PositiveEmotions	(4.333, 4.792]	3
EmoUnderstanding	(0.0914, 0.631]	2
TheoryofMind	(1.5, 1.67]	2
SocialCompParentsDec	[2, 2.429]	1

Table A.9: The frequency of each feature bin for turma9morning

feature	bin	frequency
SocialCompTeachersDec	[1.571, 2.286]	136
gender	1	92
gender	2	51
EmoRegulation	[1.125, 2]	43
CapacitytoCalmDown	(4, 5]	40
EmoUnderstanding	(-0.368, 0.0914]	32
CapacitytoCalmDown	(3.333, 3.667]	26
Empathy	(2.872, 3.337]	24
SocialCompParentsDec	(2.857, 3]	23
SocialCompParentsDec	(2.429, 2.69]	22
PositiveEmotions	(4.792, 5]	20
SocialCompParentsDec	(2.69, 2.857]	18
TheoryofMind	(1.67, 2]	14
EmoRegulation	(2, 2.375]	13
SocialCompTeachersDec	(2.429, 2.714]	13
MotorCompetence	[-1.536, -0.5]	12
Empathy	[1.8, 2.7]	12
Aggression	[1, 1.5]	11
PositiveEmotions	(3.833, 4.333]	11
TheoryofMind	(1.49, 1.5]	11
Empathy	(2.7, 2.872]	7
Aggression	(1.917, 2.167]	6
MotorCompetence	(-0.5, 0.072]	5
EmoUnderstanding	(0.0914, 0.631]	5
TheoryofMind	(1.5, 1.67]	5
MotorCompetence	(0.072, 0.437]	4
Aggression	(1.5, 1.917]	3
Aggression	(2.167, 4]	3
EmoUnderstanding	[-2.481, -0.368]	2
PositiveEmotions	(4.333, 4.792]	2
MotorCompetence	(0.437, 1.265]	1
CapacitytoCalmDown	[2.333, 3.333]	1

Table A.10: The frequency of each feature bin for turma9afternoon

feature	frequency
gender	396
TheoryofMind	214
SocialCompTeachersDec	211
CapacitytoCalmDown	178
SocialCompParentsDec	172
EmoUnderstanding	171
Empathy	159
MotorCompetence	149
PositiveEmotions	135
EmoRegulation	134
Aggression	122

Table A.11: The frequency of each feature for turma6afternoon

feature	frequency
gender	489
TheoryofMind	298
SocialCompTeachersDec	232
CapacitytoCalmDown	191
Aggression	180
EmoUnderstanding	174
SocialCompParentsDec	170
Empathy	170
EmoRegulation	168
MotorCompetence	152
PositiveEmotions	131

Table A.12: The frequency of each feature for turma6morning

feature	frequency
gender	287
Empathy	192
CapacitytoCalmDown	176
EmoRegulation	141
Aggression	139
TheoryofMind	134
PositiveEmotions	134
SocialCompTeachersDec	129
MotorCompetence	123
EmoUnderstanding	120
SocialCompParentsDec	115

Table A.13: The frequency of each feature for turma7afternoon

feature	frequency
gender	355
SocialCompTeachersDec	182
CapacitytoCalmDown	182
Empathy	181
EmoRegulation	152
TheoryofMind	152
PositiveEmotions	152
Aggression	131
MotorCompetence	123
SocialCompParentsDec	119
EmoUnderstanding	117

Table A.14: The frequency of each feature for turma7morning

feature	frequency
gender	283
EmoRegulation	147
CapacitytoCalmDown	113
TheoryofMind	111
Aggression	94
SocialCompParentsDec	82
EmoUnderstanding	75
Empathy	72
MotorCompetence	62
PositiveEmotions	60
SocialCompTeachersDec	51

Table A.15: The frequency of each feature for turma8afternoon

feature	frequency
gender	402
TheoryofMind	198
EmoRegulation	160
Empathy	150
CapacitytoCalmDown	146
SocialCompTeachersDec	145
Aggression	145
SocialCompParentsDec	105
EmoUnderstanding	94
MotorCompetence	91
PositiveEmotions	82

Table A.16: The frequency of each feature for turma8morning

feature	frequency
SocialCompTeachersDec	136
gender	111
SocialCompParentsDec	66
EmoRegulation	65
CapacitytoCalmDown	65
EmoUnderstanding	44
Empathy	31
PositiveEmotions	25
MotorCompetence	22
Aggression	21
TheoryofMind	14

Table A.17: The frequency of each feature for turma9afternoon

feature	frequency
SocialCompTeachersDec	149
gender	143
CapacitytoCalmDown	67
SocialCompParentsDec	63
EmoRegulation	56
Empathy	43
EmoUnderstanding	39
PositiveEmotions	33
TheoryofMind	30
Aggression	23
MotorCompetence	22

Table A.18: The frequency of each feature for turma9morning

feature	range ranking
Gender	0 - 1
CapacitytoCalmDown	2 - 4
Empathy	1 - 7
EmoRegulation	1 - 9
SocialCompTeachersDec	0 - 10
TheoryofMind	1 - 10
SocialCompParentsDec	2 - 10
Aggression	4 - 10
EmoUnderstanding	5 - 10
PositiveEmotions	6 - 10
MotorCompetence	7 - 10

Table A.19: Highest and lowest frequency rank for every feature across all interaction datasets

Bibliography

[Berger-Wolf and Saia(2006)] T. Y. Berger-Wolf and J. Saia, "A framework for analysis of dynamic social networks," *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '06*, 2006.

[Veiga et al.(2016)Veiga, Ketelaar, Leng, Cachucho, Kok, Knobbe, Neto, and Rieffe] G. Veiga, L. Ketelaar, W. D. Leng, R. Cachucho, J. N. Kok, A. Knobbe, C. Neto, and C. Rieffe, "Alone at the playground," *European Journal of Developmental Psychology*, p. 118, 2016.

[Rubin et al.(2009)Rubin, Coplan, and Bowker] K. H. Rubin, R. J. Coplan, and J. C. Bowker, "Social withdrawal in childhood," *Annual Review of Psychology Annu. Rev. Psychol.*, vol. 60, no. 1, p. 141171, 2009.

[Coplan and Armer(2007)] R. J. Coplan and M. Armer, "A "multitude" of solitude: A closer look at social withdrawal and nonsocial play in early childhood," *Child Dev Perspectives Child Development Perspectives*, vol. 1, no. 1, p. 2632, 2007.

[Moser et al.(2009)Moser, Colak, Rafiey, and Ester] F. Moser, R. Colak, A. Rafiey, and M. Ester, "Mining cohesive patterns from graphs with feature vectors," *Proceedings of the 2009 SIAM International Conference on Data Mining*, p. 593604, 2009.

[Aggarwal and Abdelzaher(2011)] C. C. Aggarwal and T. Abdelzaher, "Integrating sensors and social networks," in *Social network data analytics*. Springer, 2011, ch. 14, pp. 379–412.

[Akoglu et al.(2012)Akoglu, Tong, Meeder, and Faloutsos] L. Akoglu, H. Tong, B. Meeder, and C. Faloutsos, *PICS: Parameter-free identification of cohesive subgroups in large attributed graphs*, 2012, pp. 439–450.

[Zhou et al.(2009)Zhou, Cheng, and Yu] Y. Zhou, H. Cheng, and J. X. Yu, "Graph clustering based on structural/attribute similarities," *Proc. VLDB Endow. Proceedings of the VLDB Endowment*, vol. 2, no. 1, p. 718729, Jan 2009.

[Gottman and DeClaire(1997)] J. M. Gottman and J. DeClaire, *The heart of parenting: how to raise an emotionally intelligent child*. Simon Schuster, 1997.

[Zhang(2005)] B. Zhang, "Center-based clustering and regression clustering," *Encyclopedia of Data Warehousing and Mining*, p. 134140, 2005.

[Fonhof(2016)] A. Fonhof, "Patterns of children on the playground," https://www.youtube.com/playlist?list=pl_r61urvurg3ac4aqhftjy7lqoivfpfvf, 2016, accessed: 11-07-2016.

[Chakrabarti et al.(2006)Chakrabarti, Kumar, and Tomkins] D. Chakrabarti, R. Kumar, and A. Tomkins, "Evolutionary clustering," *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '06*, 2006.

[Lin et al.(2010)Lin, Sun, Cao, and Liu] Y.-R. Lin, J. Sun, N. Cao, and S. Liu, "Contextour: Contextual contour visual analysis on dynamic multi-relational clustering," *Proceedings of the 2010 SIAM International Conference on Data Mining*, p. 418429, 2010.