Received: 22 November 2020 | Accepted: 17 February 2021
DOI: 10.1111/exsy.12686

ORIGINAL ARTICLE

Expert Systems WILEY

"Want to come play with me?" Outlier subgroup discovery on spatio-temporal interactions

Carolina Centeio Jorge[1,2] | Martin Atzmueller[3] | Behzad M. Heravi[4] | Jenny L. Gibson[5] | Rosaldo J. F. Rossetti[1] | Cláudio Rebelo de Sá[2,6]

[1]Department of Informatics Engineering, University of Porto, Porto, Portugal
[2]Datamanagement & Biometrics, University of Twente, Enschede, The Netherlands
[3]Department of Cognitive Science and Artificial Intelligence, Tilburg University, Tilburg, The Netherlands
[4]Department of Computer Science, University College London, London, UK
[5]Centre for Research on Play in Education, Development and Learning, Faculty of Education, University of Cambridge, Cambridge, UK
[6]Computer Science Institute of Leiden University (LIACS), University of Leiden, Leiden, The Netherlands

Correspondence
Carolina Centeio Jorge
Email: C.Jorge@tudelft.nl

Present address
Van Mourik Broekmanweg 62628 XE Delft
The Netherlands

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: AT 88/4-1; Economic and Social Research Council, Grant/Award Number: ES/N006577/1; Kids First, Grant/Award Number: 68639

Abstract

Our lives are made of social interactions which can be recorded through personal gadgets as well as sensors capturing ubiquitous and social data. This type of data, such as spatio-temporal data from the real-time location of people, for example, can then be used for inferring interactions which can be translated into behavioural patterns. In this paper, we consider the automatic discovery of exceptional social behaviour from spatio-temporal interaction data, focusing on two areas: exceptional subgroups and spatio-temporal outliers – both in the form of descriptive patterns. For that, we propose a method for *exceptional social behaviour discovery*, combining subgroup discovery and network science methods for identifying behaviour that deviates from the norm. We also propose the use of two outlier detection metrics for identifying outliers, namely the Local Outlier Factor (LOF) and the Voronoi area. We applied the proposed method on synthetic data as well as two real datasets containing location data from children playing in the school playground. Our results indicate that this is a valid approach which is able to obtain meaningful knowledge from the data.

KEYWORDS
network science, outlier detection, play, social interactions, subgroup discovery

1 | INTRODUCTION

In today's world, a great amount of data capturing human behaviour is being collected (Terry et al., 2002). Deliberately gathering data from social and ubiquitous environments through sensors (e.g., proximity or geo-localization) is also being used to study the behaviour and interactions of people without interfering with their actions (Heravi et al., 2018). Interactions may follow patterns, sequences of behaviours, or be expressed with verbal and non-verbal gestures which we do not even notice (Goffman, 1967). Such interactions allow us to study human beings as social entities (Cabrera-Quiros et al., 2018). Then, some phenomenons may emerge (Atzmueller, 2018), such as homophily (McPherson et al., 2001), which is the tendency of people to interact more with those who are rather similar to them, etc. This suggests that socio-demographic characteristics, as well as behavioural patterns, tend to be localized (McPherson et al., 2001). In particular, there may be some local patterns which do not follow the norm, making them unusual. The automatic extraction of descriptive knowledge from the data, such as subgroups, can then help and support the analysis and decisions of social sciences experts.

Social interactions can then be observed through spatio-temporal data. In particular, by recording the sequence of people's positions over time, and their *movement data* (Lauw et al., 2010; Messinger et al., 2019; Terry et al., 2002). Then, social interactions can be analysed accordingly. On the other hand, people involved in these movement contexts, that is, the actors, have individual demographic properties that can be associated to these movements and, consequently, interactions. This being said, the analysis of social interactions requires appropriate representation and methods to study both movement data and demographic properties.

The automatic extraction of exceptional behaviour from interaction data has been already tackled in recent literature. Compositional subgroup discovery on attributed social interaction networks (Atzmueller, 2018) allows us to both represent and analyse social interactions in a spatio-temporal scenario. These proposed representation structures take into account the relative location of the people, in relation to the ones near them. The authors present different subgroup discovery quality measures that can be applied to the proposed structures. These quality measures find subgroups of people whose interactions deviate from the norm. However, this approach does not take into account the movement itself.

Another useful way of finding an observation (or a set of observations) which appears to be inconsistent with the remainder of a movement dataset is through Outlier Detection (Djenouri, Belhadi, Lin, Djenouri, & Cano, 2019). Outlier Detection has been explored in several fields (Konijn, 2017; Shekhar et al., 2003). In particular, with the massive amounts of data from urban cities, it has become very important for the study of traffic flows (Djenouri, Belhadi, Lin, & Cano, 2019; Djenouri, Belhadi, Lin, Djenouri, & Cano, 2019). Furthermore, some approaches (Liu et al., 2011) show good results when using Outlier Detection for pattern-mining algorithms, such as Subgroup Discovery. A further novel approach, proposed in this paper, is using outlier indicators throughout the sequential positions of people to complement subgroup discovery on *social interaction networks* for identifying exceptional behaviour.

In this paper, we substantially extend the work on Exceptional Behaviour Discovery (C. C. Jorge et al., 2019), with the specific target of outlier subgroup discovery on spatio-temporal (social) interactions, where we also consider the descriptive characteristics of the involved actors. The goal of this paper is thus to detect and extract characteristics of exceptional behaviour in datasets containing both movement and demographic characteristics. We define as exceptional behaviour, behaviours that are unusual or unexpected. To do so, we combine subgroup discovery techniques with network science and outlier detection techniques for the study of social interactions. To the best of the authors' knowledge, this is the first time that such an approach has been proposed.

In particular, we propose to use *digraphs* to represent *social interaction networks*, which are constructed from the movement and socio-demographic data. Furthermore, we investigate and include spatio-temporal indicators, for each individual present on the dataset, based on outlier detection methods, for example, the local outlier factor (LOF) as well as Voronoi diagrams, regarding the respective spatial distribution. To detect exceptional behaviour from this pre-processed data, we propose a method based on subgroup discovery (Klösgen, 2002), a descriptive data mining technique that provides easy-to-understand results in the form of interpretable patterns. It finds subgroups of objects (e.g., subgroups of individuals or their interactions), in movement data, that share the same characteristics with respect to a property of interest (target) (Atzmueller, 2015; Herrera et al., 2011). In this work, we consider time of interaction, time between interactions, and spatial outlierness as possible properties of interest.

Our contribution is three-fold and summarized as follows:

1. We present the novel method of *Exceptional Social Behaviour Discovery* which aims at detecting social behaviour which deviates from the norm, specifically targeting patterns of unusual behaviour in movement data – for outlier subgroup discovery.
2. In that context, we formalize these methods for analysis on social interaction networks, derived from spatio-temporal data. Specifically, we consider compositional subgroup discovery and outlier detection approaches that are combined into our proposed method for exceptional social behaviour discovery.
3. We present the results of applying our proposed method on synthetic data as well as two real-world datasets. The results demonstrate the efficacy and validity of our proposed method.

For evaluation and validation of the proposed approach, we applied it on a synthetic dataset with simple characteristics for exemplifying and demonstrating our approach. In addition, we performed two case studies on real-world data covering spatio-temporal social interactions. In particular, we tested the proposed approach on two sets of data of movement data with corresponding demographic information about the involved actors. Specifically, we utilised datasets containing locations and personal attributes of children in the school playground. The data was collected using location sensors during the school breaks. One dataset, *playgroundA* (Heravi et al., 2018), has the geographic position of 18 children over time, in 10 different sessions and personal attributes (gender, age, teacher rated social, emotional and communication difficulties). The other dataset, *playgroundB* (Messinger et al., 2019) has the position of 16 children and socio-demographic attributes, such as gender and age. The results were mostly expected, when analysed by experts (Messinger et al., 2019) in the domain. In addition, we observed similar findings and patterns when comparing between the two datasets. This indicates the validity of the approach. Furthermore, it is worth mentioning that the results added meaningful information to the expected scenario.

The remainder of this paper is structured as follows: in Section 2 we present the background, in which we explain the underlying concepts and literature review; in Section 3 we present our contributions for the state of the art, testing it with a case study presented in Section 4; we finally conclude in Section 5.

2 | PRELIMINARIES

Many domains in which we can potentially use data mining occur in a temporal or spatial scenario, providing temporal and spatial properties (Roddick & Spiliopoulou, 1999), which can provide information about objects' location, also known as *movement data* (Lauw et al., 2010). Below, we review key concepts of the techniques we use to analyse this data, that is, concepts from Subgroup Discovery, Network Science, and Outlier Detection.

2.1 | Subgroup discovery

Subgroup Discovery (SD) is a descriptive and exploratory data mining technique to identify interesting patterns, so-called subgroups, that deviate from the norm (Atzmueller, 2015; Klösgen, 2002). These patterns show an unusual distribution when compared to the overall population. This interesting behaviour is typically based on some criteria which balances their relevance between their size and atypicality. We can find SD applications in many domains, including for example, the medical (Gamberger & Lavrac, 2002), marketing (Berlanga et al., 2006), education (Romero et al., 2009), socio-demographic (Klösgen & May, 2002) and social domains (Atzmueller, 2018). As in Duivesteijn and Knobbe (2011), we define a dataset Ω as a bag of n records of the form of $x = (a_1, \ldots, a_m, t_1, \ldots, t_l)$, where a_i is a descriptor, t_i is a target, $l \in [1; m]$. Attributes are taken from a domain $\mathcal{A}$. Subgroups are usually described with a description language $\mathcal{D}$, and are induced by a *pattern*. A *pattern p* is a function $p : \mathcal{A} \rightarrow \{0,1\}$; it *covers* a record x iff $p(a_1, \ldots, a_m) = 1$. The *subgroup* induced by a pattern p is the bag of records, S_p, that p covers: $S_p = \{x \in \Omega \mid p(a_1, \ldots, a_m) = 1\}$. $p \in \mathcal{D}$ is typically a conjunction of conditions on attributes, for example, Gender = F $\wedge$ Age $\leq$ 22.

The interestingness of subgroups is measured by *quality measures* according to different types of targets. Given a subgroup discovery algorithm, a set of subgroups is identified and scored by a quality function (Klösgen, 2002): $\varphi : \mathcal{D} \rightarrow \mathbb{R}$. Quality measures are a key factor for the extraction of knowledge because the interest obtained depends directly on them (Herrera et al., 2011). Many have been proposed for identifying different deviations in different targets. Targets can be binary (Wrobel, 1997), nominal (Berlanga et al., 2006), numeric (Grosskreutz & Rüping, 2009), ranked (de Sá et al., 2016), multi-target (Atzmueller, 2015), or consider a distribution (A. M. Jorge et al., 2006).

2.2 | Network science

Network Science combines ideas from several domains of knowledge to address questions about networks (Newman, 2010). A network is a collection of *nodes* connected with *edges*. This simple representation allows one to translate many events into the form of networks, which can often lead to new and useful insights (Newman, 2010). Some key concepts of Network Science are centrality measures, which measure the nodes that are the most important or central in a network. Centrality gathers a wide range of metrics and measures that can allow us to better understand the data. For example, *degree* centrality (based on the number of edges of a node), *closeness* (based on the average length of the shortest path between the node and all other nodes in the graph), *betweenness* (based on how many shortest paths of the graph go through a node) and *pagerank* (measured by the links to a node). Some metrics proposed more recently are *hubs* and *authorities* (Kleinberg, 1999), which are defined in mutual recursion. Intuitively, we consider nodes with many outgoing links vs. nodes with many incoming links. A hub is a node with many outgoing links to authorities, whereas an authority is a node with many links from hubs. Another network concept of practical importance is provided by communities (Newman, 2010) in networks. Communities are tightly knit groups within a larger, more loosely connected network.

A particular case of networks are social interaction networks (Wasserman & Faust, 1994) which focus on interactions between people as the corresponding actors. In this case, the nodes represent the actors and the edges, the links between actors, model an interaction or event. These edges may have properties, such as frequency of occurrence or duration. Furthermore, edges and nodes may have other labels, leading to attributed networks. From these attributed networks, we can extract and characterize subgroups (Atzmueller, 2018).

A complex network can be represented by a graph (Bondy & Murty, 1976). A graph G is an ordered triple $(V(G), E(G), \psi_G)$, where $V(G)$ represents the set of vertices, $E(G)$, the edges and ψ_G is the function that associates to each edge of G a pair of vertices of $V(G)$. For example: $V(G) = \{v_1, v_2, \ldots, v_n\}$, $E(G) = \{e_1, e_2, \ldots, e_m\}$ and $\psi_G(e_1) = (v_1, v_2)$. A graph can be *directed* or *undirected*. In the case of G being directed, the output of the function $\psi_G(e_i)$, (v_j, v_k) is ordered and it is known as a *digraph* (Newman, 2010). Moreover, the graph can have multiple edges between two nodes. These graphs are referred to as *multigraph*. In particular, they can be referred to as *multidigraph* if the edges are directed. The function ψ_{MG} of a multigraph returns the same pair of vertices for more than one edge.

Some approaches combine Subgroup Discovery and Network Science. Atzmueller (2014) gave an overview of data mining in social interaction networks, specifically human behavioural (offline) networks, w.r.t. describing and characterizing networks and their properties. In terms of community detection, Skrlj et al. (2017) introduced the Community-Based Semantic Subgroup Discovery (CBSSD), an algorithm that identifies classes of instances based on structural properties of complex networks. Atzmueller (2018) also proposed quality measures and targets on interaction network properties for subgroup discovery in attributed social networks, as compositional subgroup discovery. Furthermore, Atzmueller et al. (2016) proposed an approach for description-oriented community detection using Subgroup Discovery.

2.3 | Outlier detection

Sometimes, there are points in the data that deviate from the general behaviour that are not shown in social interaction networks. These points appear to be inconsistent with the remainder, not belonging to any subgroup thus being single exceptional instances, which are also known as outliers (Barnett & Lewis, 1994; Hawkins, 1980). As such, they can also be seen as exceptional behaviour, providing special patterns with meaningful insights (Shi et al., 2018).

Breunig et al. (2000) presented a density-based approach for finding outliers in a multi-dimensional dataset. In particular, the authors present the LOF which is a possible way to measure the level of *outlierness* of each object in the dataset. Basically, the LOF reflects how close a point is to other points, translating a degree of isolation. This measure is based on a density of neighbours, reflecting the isolation in a spatial representation of the dataset. We formalize the necessary notions below as follows. Let k be a natural number:

1. The k-distance of an object o is defined as the distance, d, between o and the k_{th} closest object in the dataset D, o_k, meaning that:
   a. for at least k objects $o' \in D \backslash \{o\}$, it holds that $d(o, o') \leq d(o, o_k)$
   b. for at most $k-1$ objects $o' \in D \backslash \{o\}$, it holds that $d(o, o') < d(o, o_k)$
2. The objects o' whose distance from object o is not greater than the *k-distance* compose the *k-distance neighbourhood* $kN(o)$, $kN : D \rightarrow 2^D$, of the object o. This is equivalent to the original formalization in Breunig et al. (2000), given that the k nearest neighbours in a given neighbourhood/region (*MinPts* in the original paper – specifying a minimal number of objects) are considered.
3. The *reachability distance* $reachdist_k(o, o')$, of an object o' with respect to the object o is the k-distance if o' is part of the k-distance neighbourhood of o, otherwise it is the distance between the two objects:

$$reachdist_k(o,o') = \max\{\text{k}-\text{distance}(o), d(o,o')\}.$$

4. The *local reachability density* for a given k, lrd_k, of the object o is, intuitively, the inverse of the average reachability distance based on the k neighbours of o. It can be defined as:

$$lrd_k(o) = \frac{1}{\frac{\sum_{o' \in kN(o)} reachdist_k(o,o')}{|kN(o)|}}.$$

5. Finally, the LOF of the object o captures the degree to which we call o an outlier. It is the average of the ratio of the local reachability density of o and of the k-nearest neighbours of o. The lower local reachability density of o is, and the higher the local reachability densities of k-nearest neighbours of o are, the higher is the LOF value of o. It can be formally defined as:

$$LOF_k(o) = \frac{\sum_{o' \in kN(o)} \frac{lrd_k(o')}{lrd_k(o)}}{|kN(o)|}.$$

Thus, this method depends on the user to choose the parameter k which may be seen as a disadvantage. Qu (2008) suggest the use of Voronoi diagrams to measure the *outlierness* and avoid the parameter k. The authors define a Voronoi diagram as a subdivision of the objects into Voronoi cells. The Voronoi cell, $V(o)$ for an object o, is composed of the set of points s in the space that are closer to o than to any other object $o' \in D \backslash \{o\}$:

$$V(o) = \{s | d(o,s) \leq d(o',s), \forall o' \in D \setminus \{o\}\}.$$

Moreover, Zwilling and Wang (2014) propose Multivariate Voronoi Outlier Detection for outlier detection in multivariate time series through Voronoi diagrams which plays an important role in healthcare delivery and management domains.

Frequent pattern mining approaches for the detection of outliers have been used in urban traffic data (Djenouri, Belhadi, Lin, & Cano, 2019; Djenouri, Belhadi, Lin, Djenouri, & Cano, 2019). The data is firstly structured in patterns according to some variable(s), such as traffic volumes, congestions, and incidents. The discovered patterns are then used to find characteristics of the anomalies and their possible relationships, such as causal interaction, congested patterns, hot spot detection, and so on. In particular, Liu et al. (2011) builds a region graph from spatio-temporal urban data, detects outliers from graph edges and finally discovers relations among the outliers. This helps to identify the causality of a traffic anomaly. Pattern-mining-based approaches for outlier detection can find descriptive correlation between outliers. In fraud detection, for example, Konijn (2017) also presents a successful approach when using Subgroup Discovery with Outlier Detection. In a database of registers of a health insurance company, finding an outlier (that may mean fraud) does not give as much information as finding a subgroup of outliers. This can identify patients, pharmacies or GP's. The method uses standard techniques for measuring outlierness of single records and then looks for frequent subgroups, having the outlierness as target. The authors applied this method to identify suspicious pharmacies.

Outlier detection is then helpful to find objects that behave in an exceptional way. As previously discussed, this can be useful to detect the cases that are not described in any of the subgroups. Furthermore, we use outlier measures as targets for subgroup discovery, finding outliers with interpretable descriptions, that is, a subgroup pattern describing the outliers. In that way, we can provide more interpretable descriptions for humans, and also uncover the specific relevant factors that characterize the outliers and specifically distinguish them from the remaining set of objects.

3 | EXCEPTIONAL SOCIAL BEHAVIOUR DISCOVERY

In this paper we propose *exceptional social behaviour discovery*. The aim of the proposed method is to look for social behaviour which deviates from the norm. In order to recognize unusual social behaviour among individuals (in social interactions), we adapted an existing subgroup discovery technique to deal with spatio-temporal data in the social interactions domain. We focus on the study of subgroup discovery methods and metrics of social networks analysis and outlier detection. Part of this work extends the work proposed in (Atzmueller, 2018; C. C. Jorge et al., 2019) which combined Subgroup Discovery with social interaction networks for detecting exceptional behaviour, applying *Compositional Subgroup Discovery* on complex interaction networks. This part only considers subgroups that interact exceptionally, neglecting to detect exceptional behaviour in individuals that do not interact or fit a subgroup. Thus, we complement it with outlier detection metrics in this context, applying specific metrics as extended features in the subgroup discovery process.

3.1 | Subgroup discovery: Compositional analysis

Compositional subgroup discovery focuses on the detection of subgroups in a network, considering a *dyadic perspective*, that is, the subgroups are described by a specific pattern, while taking into account dyadic information with respect to the links/edges of the network. A dyad denotes a significant relationship between two actors of a network, which is represented by an edge connecting two nodes of the respective graph. In that sense, it differs from conventional (standard) subgroup discovery since the subgroup objects (vertices/nodes) are described by a set of attributes forming the pattern, while in our approach the quality of the subgroup is estimated on the dyads (edges). The network is then represented as a graph, where each individual is represented by a node and each interaction link contained in the network is represented by an edge between the two respective nodes. In this graph representation, both nodes and edges can be characterized by attributes, such that these can be used to find subgroups and to explain some observed behavioural patterns. For outlier detection, we can apply compositional as well as standard subgroup discovery. In the context of this paper, we specifically focus on outliers derived from spatio-temporal information using the LOF and the Voronoi area as discussed above.

To measure the interestingness, the duration of the interactions and frequency are considered. As explained in Section 2.1, we define a dataset Ω as a bag of n records given in the form of $x = (a_1, \ldots, a_m, t_1, \ldots, t_l)$, where a_i is a descriptor, t_i is a target and $l \in [1; m]$. For this problem, we consider only one target, which we name t_p, which is numeric and corresponds to the observed number of edges normalized by the expectation.

We already proposed different quality measures for compositional subgroup discovery on undirected, directed (multi-)graphs (Atzmueller, 2018; C. C. Jorge et al., 2019), which we provide for easing the discussion below.

3.1.1 | Quality measure – Simple attributed weighted graph

For the detection of subgroups in a network, considering a *dyadic perspective*, Atzmueller (2018) proposed two quality measures. The first measure uses simple attributed weighted graphs, while the second one includes frequency information in an attributed weighted multigraph representation. In all cases, the weights represent time.

In the first approach, the simple attributed graph, the weights of all the edges E_p, covered by a *pattern P*, are summed, normalized by the number of possible edges, n_E, among the nodes covered by the *pattern*, n_{Ep}. Then, r samples of n_E edges, where

$$n_E = \frac{n_{Ep}(n_{Ep}-1)}{2},$$

are considered as well as are their normalized sum of weights. Finally, a Z-score is calculated estimating the significance of the obtained value (t_P) among the samples. The Z-score gives the distance between the obtained value (t_P) and the mean value of the samples, μ_t, in units equal to the standard deviation, σ. It is calculated as $Z = \frac{t_P - \mu_t}{\sigma}$. For a *pattern p*, this quality function, q_S, is given as follows:

$$q_S(P) = Z\left(\frac{1}{n_E}\cdot\sum_{e \in E_P} w(e)\right) \tag{1}$$

3.1.2 | Quality measure – Attributed weighted multigraph

For the detection of subgroups in a network represented by a multigraph instead of a simple graph, considering a *dyadic perspective*, the authors extended the quality measure q_S. For this version, the frequency (apart from the duration) of interaction is also taken into account. Thus, for normalizing the sum of weights of a pattern p, we have to consider the multiple edges that exist between two nodes. In this case, instead of dividing by n_E, the author divides by $n_e + m_E$, where

$$m_E = \sum_{i=1}^{n_E}(m_i - 1)$$

and m_i is the observed multiplicity of an edge. Hence, for computing the Z-score, all the edges are considered.

3.2 | Compositional subgroup discovery on directed graphs

In this work, we propose to use *digraphs* to represent the interactions of the individuals. This is a direct extension of Compositional Subgroup Discovery presented in Section 3.1. To represent interactions in a directed way, we need to define (1) proximity and (2) an individual approaching another individual. If an individual approaches another within a certain proximity, a directed edge is created from the node of the individual approaching to the node of the individual approached. This approach combines *movement data* and *social data* of the subjects and returns subgroups, according to the desired quality function. The *movement data* consists of a timestamp of the event, the *id* and position (*x* and *y*) for each individual. From that, there is a function that computes the speed, *velX* and *velY*, relatively to *x* and *y*, respectively. The *social data* has the ids and socio-demographic data corresponding to the individuals in the *movement data*. Any numeric attributes are discretized in equal frequency bins.

We also propose two variants to weight the digraphs and multidigraphs. In the first variant, the weight represents the duration of interaction. In the simple attributed digraph, this is the sum of all interactions. Moreover, in the attributed multidigraph the weight of each edge is the time of that interaction. The second variant takes into account the time between interactions. Meaning that in the simple attributed digraph, the weight of an edge is the sum of the times between interactions. Furthermore, in the attributed multidigraph the weight of each edge is the time between the end of the previous interaction (an edge only exists if there has been interaction before) and the start of the next one.

3.2.1 | Generating the interaction digraph

To create the interaction digraph, or multidigraph, we first need to define interactions. We consider an interaction between two individuals when their relative distance is within a certain *proximity* and one of the individuals approaches the other. Therefore, given a maximum distance threshold between individuals, *maxdist*, we start with an empty digraph *G*. Then, we create new edges and calculate their weights over time.

1. **Creating a new edge:** At each time step t, a matrix of distances, D, between every two individuals is calculated. Then, for each distance $d_{i,j} \in D : d_{i,j} \leq maxdist$ we compute a vector from i to j as $\vec{r}_{i,j} = (x_j - x_i, y_j - y_i)$. We then verify the speed vector of i, $\vec{v}el_i = (velX_i, velY_i)$, and calculate the cosine between the vectors $\vec{r}_{i,j}$ and $\vec{v}el_i$. If the cosine is not negative, we consider that the individual i *approached* (or *reached*)

individual *j*. Thus, we create a directed edge from node *i* to node *j* and add it to *G*. For the multidigraph version, a directed edge is added to *G* at moment *t* if individual *i* approached individual *j*, given that it was not interacting in $t-1$,

2. **Assigning a weight to the new edge:** The edge weight can represent the duration of an interaction or the time between two subsequent interactions (between the same two individuals).
   a. If the weights of edges represent the duration of interactions, for the simple digraph, when a new edge is created (from node *i* to node *j*), $w_{i,j} \in W$ is incremented by one unit of time. *W* is the matrix of weights and $w_{i,j}$ is the number of times that the individual *i* approaches the individual *j*. For the multidigraph version, $w_{i,j}$ is the total time that the individual *i* approaches the individual *j* without interruption.
   b. If the weights of edges represent the time between interactions, we use an auxiliary matrix *A*, where $a_{i,j}$ is either 1, 0, or ∞ meaning that individual *i* is interacting with individual *j*, is not interacting with individual *j*, or no information is available, respectively. The values of $a_{i,j}$ are updated at each *t* after the first time *i* approaches *j*. In the simple digraph, when a directed edge (from node *i* to node *j*) is added to *G*, $w_{i,j} \in W$ is incremented by the difference of time for the end of the last interaction every time individual *i* approached individual *j* and $a_{i,j}$ is 0. For the multigraph version, $w_{i,j}$ is the difference between the end of the last interaction's time and *t*.

3.2.2 | Quality measures

For analyzing interaction graphs, we propose two quality measures with two variations, making use of the fact that we apply a directed graph. For analyzing outliers, we apply the quality functions as discussed below.

Simple attributed digraph: q_{SD}

This quality measure takes into account the duration of the interactions between two individuals. A new directed edge (or arrow) is considered every time an interaction is observed and not clear. For the quality function, we use the same measure as q_S [Equation (1)]. However, since we have the double of the edges (because this is a directed version), we use $n_E = n_{Ep}(n_{Ep} - 1)$.

Simple attributed multidigraph: q_{SM}

This quality measure considers both the duration and frequency of the interactions between two individuals. In this case, one directed edge is created every time an interaction starts. For the quality function we proceed analogously as discussed in Section 3.1.2.

To-node and from-node variants

These two variants, *To-node* and *From-node*, extend the quality measures mentioned above. In the *To-node* and *From-node* variants, the attributes of the edges are only based on the attributes of the *head* node or the *tail* node, respectively. With these variants we hope to find valuable information about the attributes of the individuals that look for interactions (*From-node*) and the individuals that are reached the most (*To-node*).

3.2.3 | Applying subgroup discovery on compositional digraphs

The edges of the graph *G* are associated with features, which are based on the attributes of the nodes of that edge. In the comparison versions (Simple attributed digraph and Directed attributed multidigraph), numeric attributes in the nodes are compared (for each edge between them), resulting in *equal* (or same), *greater* or *lower*, which will be the features of the edges. In the *To-node* and *From-node* variants, the numeric attributes of the nodes are represented as *medium*, *high* or *low*, depending on their frequency bin, and the edge takes the features of the node *to* or *from*, respectively. Figure 1 gives an example of a Compositional Digraph and possible features of the edges.

After assigning attributes to the edges, we apply an adaptation of the SD-Map* algorithm (Lemmerich et al., 2016) on the graph. The output is a list of subgroups and their characteristics, namely pattern description, the number of edges and nodes covered by the pattern, the mean weight of those edges and the score (quality function result).

We also propose to add automatically generated features to the nodes' attributes and, consequently, to the edges' with the use of complex networks' metrics. We can add, for example, degree (also in-degree and out-degree), centrality measures (eigenvector, closeness, betweenness), authority and hub values. These metrics are based on the overall interactions, and provide more context for the interpretation of the specific subgroup.

3.3 | Outlier subgroup discovery

Subgroup Discovery looks for subgroups that differ from the overall population. In Exceptional Behaviour Discovery we aim to find unusual behaviour with a focus on human behaviour. Being lonely is typically not considered a standard human behaviour. While some people may feel

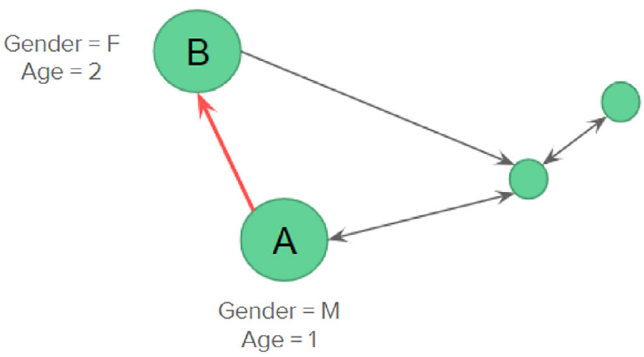


FIGURE 1 The directed edge from A to B, in red, has the following properties: {Gender = (M,F), Age = >} for simple version, {Gender = M, Age = 1} for from-node version and {Gender = F, Age = 2} for to-node version. For the sake of interpretability, we describe the subgroups (of edges) with these attributes as {Gender = A → Gender = M ∧ Age = lower → Age = higher} in the simple version

happy to spend time alone, it can sometimes be a sign of unknown issues with the peers or a particular group of people. Children in particular might be left aside due to their differences. For this reason, understanding which characteristics are more associated with the loneliness of children can support its mitigation and awareness.

Therefore, we propose two approaches to find subgroups with unusual lonely scores, using measures of outlier detection. For a database D' we consider the trajectories of individuals, that is, their positions over time. Each row of the database corresponds to a specific position (for a given time *t*) of an individual together with the associated features. For this database, we add the outlier score as an additional attribute. To this new database D' we can apply Subgroup Discovery techniques to find one or more *id* that show unusual behaviour (unusual outlier scores, which is our *target*).

We consider two different measures for computing the outlier score:

1. *Local Outlier Factor (LOF)*: Using the LOF, we can measure the isolation of a point, considering its position and its neighbours' positions. Based on that, it assigns a numeric score to each point. At each time *t* we calculate the LOF of all the objects present (that have a registered position at time *t*) and add all their attributes, along with a new numeric attribute *lof* to D'. We used *LocalOutlierFactor* from *scikit learn* (Pedregosa et al., 2011) to calculate *LOF*.
2. *Voronoi area*: Similarly, we calculate the Voronoi areas as a score for *outlierness*. The Voronoi area is the area of the corresponding Voronoi cell. We start by limiting the space based on the minimum and maximum *x* and *y* values of the individuals' positions, adding a threshold derived from a *factor* to avoid infinite values of area. This will then create a rectangular shape that contains all the objects in a snapshot. At each snapshot we calculate the Voronoi cells, their areas, and create a new instance of D' for each object gathering the attributes and the new numeric attribute *area*. For this, we used *MultiPoint*, *Point* and *Polygon* from *Shapely*[1] and *Voronoi* from *SciPy* (Jones et al., 2001).

For the search of subgroups, we consider two quality measures: one based on standard quality measures and an adapted one implementing ideas proposed for compositional subgroup discovery, as discussed above. With the first one, we show how we can use any subgroup discovery algorithm already implemented for this problem, while the second one requires specific extensions for estimating the quality. In particular, the *Z-Score-Adjusted Approach* also considers the distribution of means of random subgroups, for estimating statistical significance.

1. *Classical Approach*: For the first alternative, we can apply a standard subgroup discovery approach (e.g., SD-Map*, Lemmerich et al., 2016, or beam search, Klösgen, 2002), with a numeric target (the outlier score) and a standard quality function for numeric targets. This quality function scores a subgroup based on its unusualness (difference of the value of the target for a subgroup, *mean_sg* and the overall population, *mean_dataset*) and its frequency:

$$q_{SN} = instances\_subgroup * (mean\_sg - mean\_dataset).$$

For our experiments, we used the library *pysubgroup*, presented in Lemmerich and Becker (2018).

2. *Z-Score-Adjusted Approach*: This quality measure is a Z-score of the mean target of the under-evaluation subgroup when considering the means of random subgroups. For example, we compute the mean target of a potential subgroup, compute *r* random subgroup and respective mean targets and finally compute the Z-score of the subgroup. To generate the potential subgroups we can use a standard subgroup discovery algorithm, for example, SD-Map for generating according subgroup patterns. In our implementation, we used an according variant.

4 | A CASE STUDY IN PLAYGROUND SOCIAL INTERACTIONS

Analyzing social interactions in the playground can be of utmost importance. Social group structure and dynamics are believed to be strongly related to the child's well-being and yet has been poorly understood and studied (Heravi et al., 2018). For this reason, we used datasets of children's movement in the playground as a case study.

4.1 | Data

To test the approach explained in Section 3.2, we used two datasets with locations of children in the school playground. The data was collected with the use of location sensors during the school breaks. Furthermore, we applied a synthetically generated artificial dataset with simple characteristics for demonstrating our approach.

4.1.1 | Dataset *playgroundA*

The dataset *playgroundA* (Heravi et al., 2018) has the geographic position of 18 7–8 year-old children (9 girls) over time, during approximately 45 min. It also includes the personal attributes (gender, age, emotional stability, etc). The children were playing outdoors, without toys, during a normal day of Primary School. They had a head-mounted sensor with IMU and GNSS for precise positioning a shoe-mounted IMU sensor for activity monitoring. The following social and psychological measures were collected from a teacher:

1. **Social skills**: low, medium, high
2. **Conduct**: a high value represents behaviour problems
3. **Emotion**: the higher the score the more emotional difficulties
4. **Peer**: high scores indicate that the child has issues making friends
5. **Hyper**: the higher the more hyperactive the child is

4.1.2 | Dataset *playgroundB*

The other dataset, *playgroundB* (Messinger et al., 2019) has the position of 14 children (8 girls) around 5 years-old and socio-demographic attributes, such as gender and age. The data was collected through a real time location system that used UWB sensors. The data was collected during 1 h.

4.1.3 | Preprocessing of *playgroundA* and *playgroundB*

Both datasets present ids and positions in two dimensions. For each position of each id, we compute the instant speed, based on the next position available. For the numeric *social data* of *playgroundA*, we transformed the numeric values in 3 bins (0, 1, and 2 or *low*, *medium* and *high*). Then, we created the graphs and experimented the 6 approaches for each: comparison, to-node and from-node attributed edges for both simple (q_{SD}) and multidigraph (q_{SM}) quality measures. Furthermore, we also looked for subgroups based on Network Science metrics in the *playgroundB* dataset.

4.1.4 | Synthetic data – *Random walks* dataset

Furthermore, we created an artificial dataset, we call it *random walks*, to better evaluate our model. This artificial dataset is composed of 16 individuals' movement for 20 min, in a square space of 10×10 m. These individuals have 3 social attributes: *Att1*, *Att2*, *Att3*. The movement and social attributes of the individuals is determined as follows:

1. 10 individuals (ids from 0 to 9) start at position (0,0) and move randomly 0.2 in one of 4 directions. These individuals do not share social attributes with any other.

2. 3 individuals (ids 10, 11 and 12) stay in the center of the square for the whole time, with positions close to each other and share all social attributes with and only with each other (all three attributes have the value *a* for these individuals).
3. 3 individuals (ids 13, 14 and 15) stay at 3 different corners of the square for the whole time and share all social attributes with and only with each other (all three attributes have the value *o* for these individuals).

4.2 | Results and discussion

In this section, we analyse some of the results of our approach with *playgroundA*, *playgroundB* and *random walks* datasets. An adapted version of SD-Map* (Lemmerich et al., 2016) is used for subgroup discovery with quality measures q_{SD} and q_{SM}, in the social interactions and the *pysubgroup* python package Lemmerich and Becker (2018) for the classical approach on outlier detection. We compare some of the results with the Cortana[2] Data Mining tool for discovering local patterns in data (Meeng & Knobbe, 2011).

4.2.1 | *Random walks*: Testing the approach

We created the *random walks* dataset with a desired output in mind, as a ground truth for estimating whether our algorithms work correctly, fulfilling the expected behaviour. In particular, it is expected that the most interesting subgroups from interactions are composed of the social attributes' values of the individuals 10, 11 and 12, that stayed near each other the whole time. Thus, combinations of attributes with value *a* are expected. On the other hand, we expected we could complete this analysis by detecting the subgroups of outliers (the 3 individuals by the corners, with ids 13, 14 and 15). In this case, combinations of attributes with value *o*, and the ids 13, 14 or 15 are expected to appear as the most interesting subgroups for the outlier techniques.

Table 1 shows the results of q_{SD} for comp, to-node and from-node attributed digraph versions, defined as *comp*, *to* and *from*, in column *V*, respectively, on *random walks* dataset. These results are ranked (position showed in column *Rank* from the most interesting subgroup to the least interesting, according to its quality score, *Z*. As expected, the 7 most interesting subgroups (first in the rank) are combinations of social attributes' with values *a*, with score 42. As shown in the table, these subgroups are composed of 3 nodes (*N*) and 6 edges (*E*) with mean weight ($|C|$) 1141. These 1141, in seconds, approximately represent the 20 minutes of interaction, the whole duration of the dataset. These patterns combine both a high E and $|C|$, as expected. After these first 7 subgroups, the interestingness drops from 42 to 1 or 0, showing a big gap between the quality of the first 7 and the rest of the considered subgroups.

In Table 2, we present the results of the Z-Score-Adjusted approach for Outlier Detection subgroups, using LOF as metric. As expected, we can see that the best ranked subgroups are combinations of social attributes with value *o*. In particular, the algorithm identifies the individual 14 in the first three more interesting subgroups, which is what we expect, because the top subgroups should be composed of attributes' values *o* or the ids 13, 14 or 15. When a pattern of a top subgroup includes an id, we can interpret this as this individual being an outlier. Furthermore, the 4th subgroup describes only Att3, meaning that it found common characteristics in individuals that show outlier behaviour.

TABLE 1 Ranking of subgroups (comp, to-node and from-node attributed [simple] digraph versions) according to the total duration of interactions between every two individuals in the dataset *random walks*

Rank	V	Pattern	*N*	*E*	$\|C\|$	*Z*
1	comp	Att1 = a → Att1 = a ∧ Att2 = a → Att2 = a ∧ Att3 = a → Att3 = a	3	6	1141	42
2	comp	Att1 = a → Att1 = a ∧ Att2 = a → Att2 = a	3	6	1141	42
…	…	…	…	…	…	…
8	comp	Att1 = c → Att1 = k	2	1	114	1
1	to	Att1 = a ∧ Att2 = a ∧ Att3 = a	11	29	67.4	8.5
2	to	Att1 = a ∧ Att2 = a	11	29	67.4	8.5
…	…	…	…	…	…	…
8	to	Att1 = o ∧ Att2 = o ∧ Att3 = o	2	1	21	−0.1
1	from	Att1 = a ∧ Att2 = a ∧ Att3 = a	13	34	50.8	5.5
…	…	…	…	…	…	…
8	from	Att1 = o ∧ Att3 = o	2	1	35.5	0.1

TABLE 2 Ranking of subgroups in the Z-Score-Adjusted approach according to the LOF of each individual in the dataset *random walks*

Rank	Pattern	Mean LOF	Z
1	Att2 = o ∧ Att3 = o ∧ id = 14	4.6	10.7
2	Att1 = o ∧ id = 14	4.6	10.7
3	Att2 = o ∧ id = 14	4.6	10.6
4	Att3 = o	3.7	10.6

TABLE 3 Ranking of subgroups (comp, to-node and from-node attributed multidigraph versions) according to the total duration of interactions between every two children in the dataset *playgroundA*

Rank	V	Pattern	*N*	*E*	|*C*|	*Z*
1	comp	Gender = M → Gender = M ∧ Age = same → Age = same ∧ Social_Skills = same → Social_Skills = same	5	176	3.1	195.3
2	comp	Gender = M → Gender = M ∧ Emotion = same → Emotion = same ∧ Social_Skills = same → Social_Skills = same	5	162	2.9	179.1
3	comp	Age = same → Age = same ∧ Emotion = same → Emotion = same ∧ Social_Skills = same → Social_Skills = same	6	110	2.1	153.7
4	comp	Age = same → Age = same ∧ Emotion = same → Emotion = same ∧ Social_Skills = same → Social_Skills = same ∧ Hyper = same → Hyper = same	6	110	2.1	137.7
5	comp	Gender = F → Gender = F ∧ Emotion = same → Emotion = same ∧ Hyper = same → Hyper = same	4	287	2.7	116.8
1	to	Age = medium ∧ Gender = F ∧ Conduct = high ∧ Social_Skills = high ∧ Emotion = high ∧ Peer = medium	12	156	0.5	21.4
2	to	Conduct = high ∧ Gender = F ∧ Emotion = high ∧ Hyper = low ∧ Peer = medium	12	156	0.5	21.4
3	to	Conduct = high ∧ Social_Skills = high ∧ Emotion = high ∧ Peer = medium	12	156	0.5	21.4
4	to	Conduct = high ∧ Social_Skills = high ∧ Emotion = high ∧ Peer = medium ∧ Hyper = low	12	156	0.5	21.4
1	from	Conduct = high ∧ Social_Skills = high	12	174	0.4	14.5
2	from	Conduct = high ∧ Gender = F ∧ Emotion = high ∧ Hyper = low ∧ Peer = medium	12	174	0.4	13.7
3	from	Age = medium ∧ Gender = F ∧ Conduct = high ∧ Social_Skills = high ∧ Emotion = high ∧ Peer = medium	12	174	0.4	13.7
4	from	Age = medium ∧ Gender = F ∧ Conduct = high ∧ Social_Skills = high ∧ Emotion = high ∧ Peer = medium ∧ Hyper = low	12	174	0.4	13.7

4.2.2 | *playgroundA*: Compositional Analysis

Table 3 shows the ranked subgroups found in the dataset *playgroundA*, with quality measure q_{SM}. We present three versions: a comparison version, a to-node and from-node version with attributed multidigraph version (comp, to and from in the column V, in the Table 3, respectively). For each subgroup, we show its pattern, number of nodes (children) belonging to the subgroup, *N*, number of edges (interactions), *E*, the mean time of interactions between children in the subgroup, $|C|$, and the Z-score based on the comparison between the total duration of the interactions in the subgroup and the null model, *Z*.

The top ranked subgroups (Table 3) obtained with the comparison version show many comparison of social attributes as *same*, such as *Gender = M → Gender = M*, *Gender = F → Gender = F*, *Age = same → Age = same* and *Social_Skills = same → Social_Skills = same*. This means that children who share the same social attributes' values interact and follow each other much more than what would be expected by random interactions. We note that these subgroups have much higher scores than the others, which goes in line with the observations already discussed in Messinger et al. (2019). This seems to confirm the homophily regarding several attributes, in particular regarding the gender, meaning that children interact more with children of the same gender.

The *to-node* and *from-node* versions show very similar characteristics among each other. In fact, the top-4 ranked subgroups for both versions are very similar between each other. After analysing the number of in-nodes and out-nodes of the subgraphs covered by these patterns, we can see that the *to-node* version only has one in-node (number of nodes with positive in-degree), meaning that all interactions are towards the same child. In all subgroups, this child is the same. When checking the out-nodes (number of nodes with positive out-degree) of the *from-node* version, we also noticed that it was one in all of them and always the same one. This means that all interactions represented by the edges of these

TABLE 4 Ranking of subgroups (comp, to-node and from-node attributed [simple] digraph versions) according to the total duration of interactions between every two children in the dataset *playgroundA*

Rank	V	Pattern	*N*	*E*	|*C*|	*Z*
1	comp	Gender = F → Gender = F ∧ Emotion = same → Emotion = same ∧ Hyper = same → Hyper = same	4	12	64.1	42.8
2	comp	Gender = M → Gender = M	8	39	36.3	33.9
3	comp	Gender = F → Gender = F	8	54	35.6	32
4	comp	Gender = M → Gender = M ∧ Emotion = same → Emotion = same	6	19	31.1	28.1
5	comp	Gender = M → Gender = M ∧ Peer = same → Peer = same	8	25	28.6	27.2
1	to	Peer = low	16	122	18	5.2
2	to	Gender = M ∧ Peer = low	16	67	11.6	5.6
3	to	Gender = M ∧ Social_Skills = medium	13	21	7.5	5.1
4	to	Gender = M	16	89	13.5	5.1
1	from	Peer = low	16	123	18	6.2
2	from	Gender = M ∧ Peer = low	16	70	11.4	5.1
3	from	Conduct = low ∧ Social_Skills = medium ∧ Peer = low	13	21	7.2	4.9
4	from	Conduct = low ∧ Social_Skills = medium	13	21	7.2	4.8

subgraphs start in the same child. Furthermore, the child is the same in both versions. This shows a particular case of a child whose social interactions with others, when both duration of interaction and frequency considered, are highly uncommon and unexpected.

For a better analysis of social interactions subgroups in dataset *playgroundA*, Table 4 presents comparison, to-node and from-node versions of the quality measure for attributed directed graphs, q_{SD}. Unlike q_{SM}, q_{SD} does not take into account the frequency of interactions. Instead, it takes into account only the total duration of interaction between every two children, in a directed way. The results of the *comp* version are similar to the multidigraph version. The best ranked subgroups also show homophily, especially regarding gender. Other attributes that seem to be significant when shared between children are *Emotion*, *Hyper* and *Peer*. In particular, for the *to/from* version, it is observed that boys with *Peer = low*, meaning they present a better quality in peer relation, both reach and are reached for interactions unexpectedly – compared to random chance. It is interesting to also notice that children with *Conduct = low*, meaning that they do not show behavioural problems, and average social skills look more for interactions than what is generally observed in our study.

Results of the to-node and from-node versions can add valuable information to the results found in the comparison version, as to know what attributes the children that start more interactions or are more reached via other children have in common. For these results, we conclude that the multidigraph version presents more detailed results than the digraph version. A combination of the comparison, to-node and from-node versions with both q_{SM} and q_{SD} quality measures for attributed multidigraph and digraph representations of the interactions, can reveal very interesting and unexpected patterns in social interactions.

Table 5 shows the results when *time between interactions* was used as weight with simple attributed digraph version. The comparison (comp) versions of both approaches show similar results, presenting homophily, including age attribute. However, in the to-version *Gender = M* appears as a top subgroup, whereas in the from-version, *Gender = F*. This can be interpreted as boys being reached out after a longer time since the previous interactions, whereas girls are the ones that take more time between interactions.

4.2.3 | *playgroundA*: Outlier Analysis

In this section we analyse the Voronoi areas and LOF of the children and also present the subgroups with the most unusual scores. The subgroups identify the characteristics of the children which spent an unusual amount of time alone when most of the children spend their time in groups. Likewise, we can also expect to find subgroups indicating characteristics of children that spend most of the time near others, if the majority spends the time alone.

Finding subgroups with unusual Voronoi areas

In Figure 2 we can observe the Voronoi areas per child along the time. Each line represents one particular child's area and the green thick line is the mean area of all children throughout the time. This, combined with the demographics and social metrics represents the input data of our method.

If we consider the average Voronoi areas per child, we can see that some children show unusual values, standing out of the others (Figure 3). For example, the child with the id 17 shows the highest value and the id 2 the lowest. To verify this measure of child 17, we took snapshots of

TABLE 5 Ranking of subgroups (comp, to-node and from-node attributed simple attributed digraph versions) according to the time between interactions in the dataset *playgroundA*

Rank	V	Pattern	N	*in*	*out*	E	|B|	Z
1	comp	Gender = F → Gender = F	8	8	8	49	804.9	26.7
2	comp	Gender = M → Gender = M ∧ Social_Skills = same → Social_Skills = same ∧ Age = same → Age = same	5	5	5	16	1027	22.3
3	comp	Gender = M → Gender = M	8	8	8	33	668.6	22.3
1	to	Peer = low	16	10	16	106	414.4	6
2	to	Gender = M	16	8	16	78	335.9	5.8
3	to	Age = low	16	7	16	71	291.2	5.1
4	to	Gender = M ∧ Peer = low	16	6	16	59	261.3	5.1
5	to	Hyper = high	16	5	16	52	242.3	4.8
1	from	Gender = F	16	16	8	94	367.6	5.4
2	from	Emotion = medium	16	16	4	47	235	4.9
3	from	Peer = medium	16	16	5	54	245.3	4.8
4	from	Peer = low	16	16	10	109	378.2	4.6
5	from	Age = low	16	16	5	55	237.8	4.4

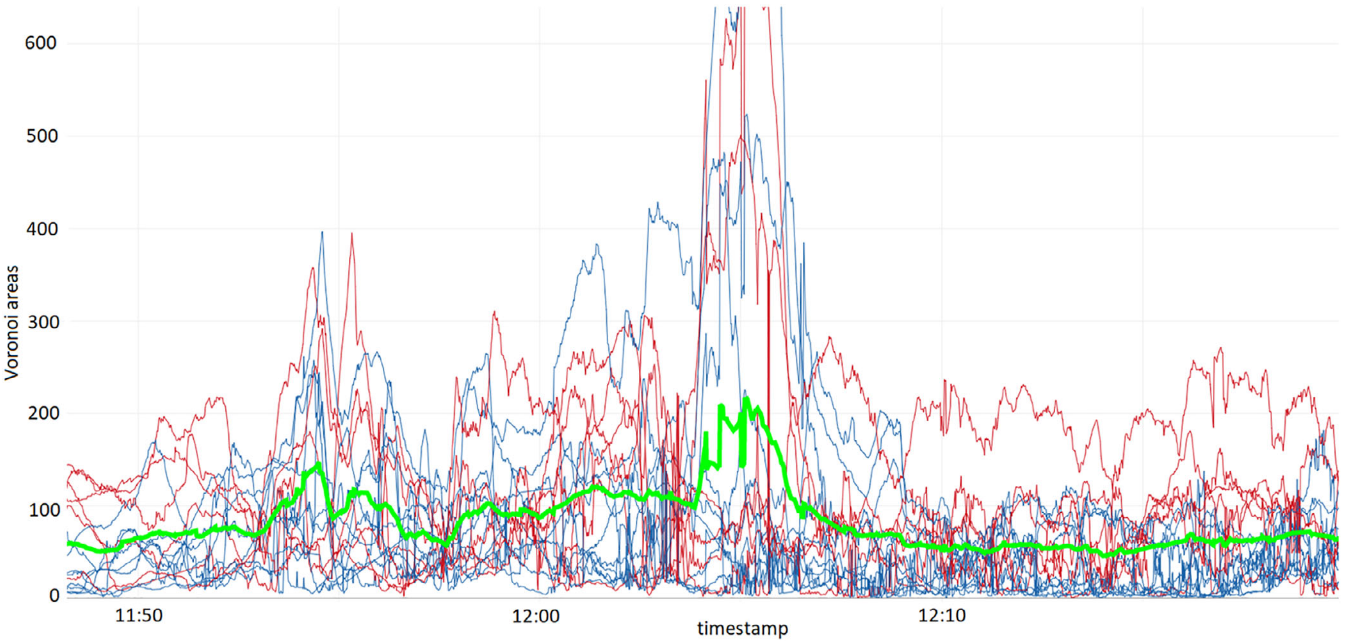


FIGURE 2 Voronoi areas of each child (per line) along time. Most scored subgroup for Z-Score-Adjusted approach with Voronoi area as target in red. (The image was cropped in height for the sake of space)

the playground position of every child in different moments in time. Some of these snapshots are presented in Figure 4 where the Voronoi areas per child are coloured. Child with id 17 (depicted by the red dot), can be seen in the peripheries of the group of children and is never found in a central position. Central positions would be closer to other children, which would be more likely to result in smaller Voronoi areas.

Finally, using the Voronoi areas as our target, we performed subgroup discovery using the Z-Score-Adjusted approach. The results are summarized in Table 6. The best and second best subgroup, "*Gender = M ∧ Social skills = Low*" and "*Social skills = Low*", represent the same group of children. They include the ids 3, 16, 18, 9 and 13 which also showed unusually high average Voronoi areas (See Figure 3). The average Voronoi area of this subgroup is 97.5, which is much higher than the average, 82.9. This finding seems to make sense since we would expect that children with lower social skills would spend less time near other children.

The 3rd and 4th subgroup also represent the same 5 children with the ids 3, 9, 13, 16 and 18. In this subgroup, 4 out of 5 children have an average Voronoi areas above the mean. As in the best two subgroups, these two also indicate low social skills as a common characteristic to have higher Voronoi areas. These are strong indicators that this method successfully extracts characteristics that accurately describe outliers.

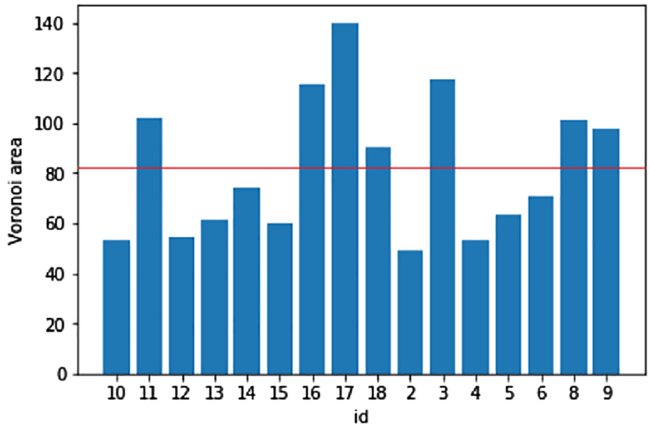


FIGURE 3 Mean Voronoi area of each child. The red line shows the average

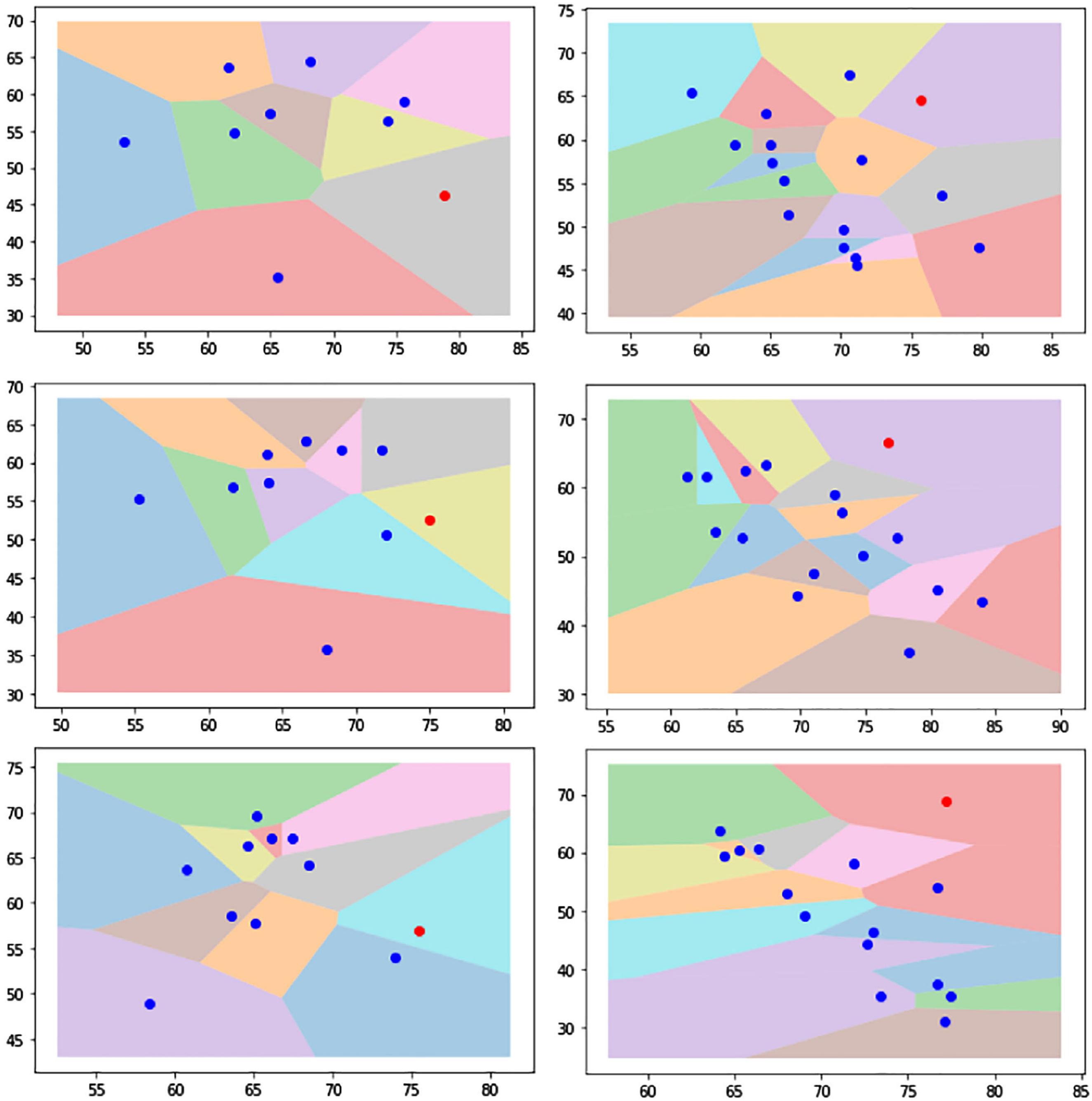
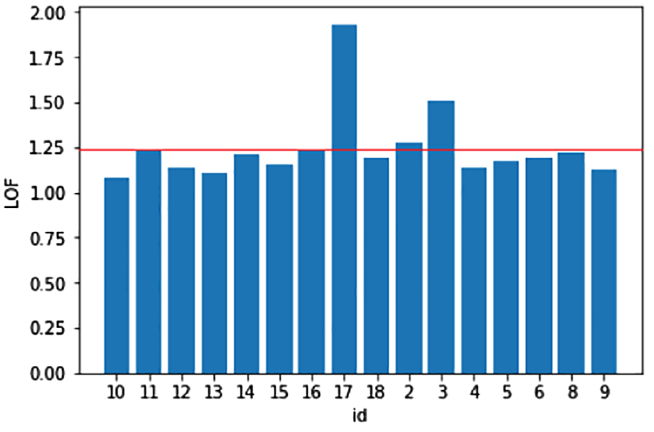


FIGURE 4 Voronoi area of each child in different moments in time. Child 17 is represented with the red dot

TABLE 6 Ranking of subgroups in the Z-Score-Adjusted approach according to the Voronoi area of each child in the dataset *playgroundA*

Rank	Pattern	Mean Voronoi	Score
1	Gender = M ∧ Social_skills = Low	97.5	32.5
2	Social_skills = Low	97.5	27.2
3	Social_skills = Low ∧ Emotion = 0	96.7	25.7
4	Gender = M ∧ Social_skills = Low ∧ Emotion = 0	96.7	25.1
5	Age = Low ∧ Social_skills = Low	93.6	21.2
6	Gender = M ∧ Social_skills = Low ∧ Age = Low	93.6	19.5

FIGURE 5 Mean LOF of each child. The red line shows the average



The classical subgroups approach, for which we used *pysubgroup* (Lemmerich & Becker, 2018) library for a beam search, using the same target, Voronoi area, found in this dataset a very similar set of subgroups. The top 3 subgroups are exactly the same as the ones in Table 6. The 4th best subgroup was "*Gender = M* ∧ *Age = 0*" which provides some new insights. These subgroups represent males in the lower group age.

Finding subgroups with unusual Local Outlier Factor scores

In this section, instead of using the Voronoi area we use the LOF measure to detect unusual behaviour. Figure 5 shows the mean LOF for each child and the average is depicted with the red line. In this case, only 2 children differ from the overall population, regarding this measure, namely 17 and 3. We note that these two ids, 17 and 3, were also very high in the plot in Figure 3. In this case, these 2 children are the only ones that stand out while all the others have scores really close to the average. In comparison to the Voronoi area, it seems that the LOF scores vary much less.

With the classical subgroup approach, with LOF as target, Table 7 shows the discovered subgroups. The parameter k, in this experiment, does not affect the results much. We opted by using $k = 3$. The top-4 subgroups had the same score and they seem to be quite different from the ones in Table 6. These subgroups include the child with id 17 (the child with highest average LOF, presented in Figure 5). Moreover, the 5th most interesting subgroup includes *id = 17*. Since the child with id 17 shows the highest LOF and there is not much variance in the LOF values, this result was very expected. The best 5 subgroups resulting from the *Z-Score-Adjusted Approach* with the LOF as target all had a Z-score = 1.3. For this reason, the findings are not considered meaningful and are not discussed in this paper. This can be due to the small variance of the LOF scores (Figure 5).

We can conclude that using Outlier Detection measures for Subgroup Discovery is a very interesting approach. In particular, Voronoi areas seem to tell more than LOF, since the Voronois areas showed more variance than LOF values. The discovered subgroups then provide characteristics of children that are considered outliers. This may be of utmost importance to identify children with higher risks of unusual behaviour. Since the id was also used as an attribute of each child, it would be expected that the algorithms found one child as the subgroup, for example, a subgroup covered by the pattern "*id = 17*", if there was only one outlier or the "most" outlier did not share characteristics with others with similar behaviour.

Finding subgroups by coordinates with Cortana

For comparison, we ran Cortana on the *playgroundA* dataset. We used the position coordinates of each child as the target, instead of pre-calculated outlier measures, such as LOF or Voronoi areas. Table 8 shows the found subgroups. The first three subgroups cover the same children (ids 3 and 13) and can be better observed in Figure 6. By looking at Figure 6, we can understand that this subgroup was found due to its unexpected coordinates near the borders of the playground. The last two subgroups present similar spatial unexpected characteristics and cover children 8 and 11.

TABLE 7 Ranking of subgroups in the classical approach with *pysubgroup* approach, according to the LOF of each child in the dataset *playgroundA*

Rank	Pattern	Score
1	Conduct = low ∧ Emotion = high ∧ Social_Skills = high	235.9
2	Conduct = low ∧ Emotion = high	235.9
3	Conduct = low ∧ Emotion = high ∧ Gender = F	235.9
4	Conduct = low ∧ Emotion = high ∧ Hyper = low	235.9
5	Peer = medium ∧ id = 17	230.4

TABLE 8 Ranking of subgroups found by Cortana, having position coordinates of each child as target

Rank	Pattern	Score
1	Peer = low ∧ Conduct = medium ∧ Gender = M	0.25929
2	Peer = low ∧ Conduct = medium ∧ Emotion = low	0.25929
3	Peer = low ∧ Conduct = medium ∧ Social_Skills = low	0.25929
4	Conduct = high ∧ Hyper = medium	0.25918
5	Conduct = high ∧ Gender = M	0.2558

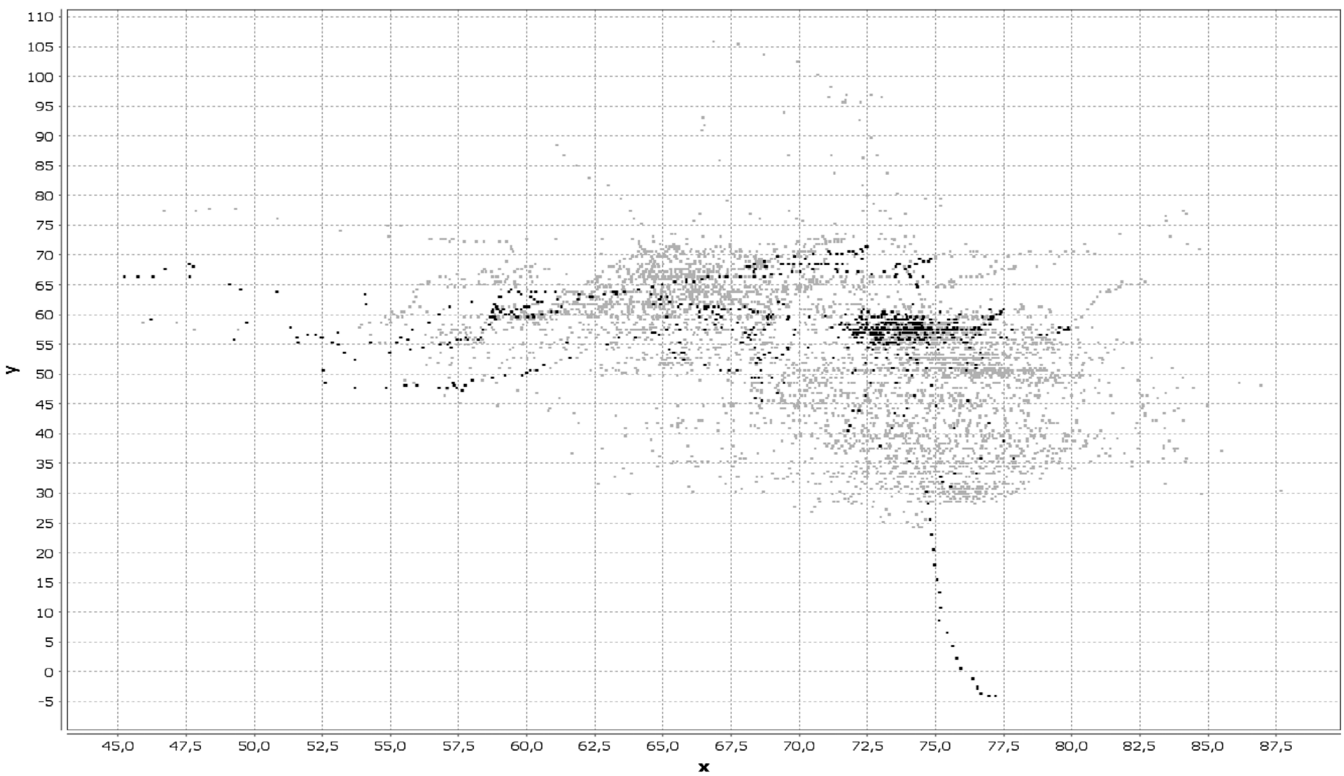


FIGURE 6 Spatial positions of children given in terms of *x* and *y* coordinates where each dot represents one child at a given point in time. The darker dots correspond to the positions of the children covered by the first subgroup presented in Table 8, Peer = low ∧ Conduct = medium ∧ Gender = M

This method shows interesting results but fails to make use of spatio-temporal properties of this dataset. While it is good at finding subgroups which contain children that were outside the expected positions, when all positions (over time) are considered, it does not compare these positions with the other children's, on different time frames. For this reason, this method may fail at finding children which share the same positions

TABLE 9 Top-4 subgroups (comparison version of simple attributed digraph) according to the total duration of interactions between every two children, considering Network Science metrics in the dataset *playgroundB*

Rank	Pattern	*N*	*E*	|*C*|	*Z*
1	Gender = M → Gender = M	6	25	0.2	21.7
2	Gender = M → Gender = M ∧ hubs = same → hubs = same	4	12	0.2	11.7
3	Gender = F → Gender = F ∧ closeness = same → closeness = same	6	12	0.1	7.7
4	Gender = F → Gender = F	8	22	0.1	7.0

with the others, but not at the same time frames. On the other hand, our methods calculate LOF and Voronoi areas, which compare a child position with other children's positions, at different time frames. Therefore, Cortana fails to identify some interesting subgroup patterns regarding this important and interesting information, and thus our method provides a more comprehensive perspective on the spatio-temporal properties.

4.2.4 | *playgroundB*

For the dataset *playgroundB*, we analyse the attributed multidigraph approach. When analyzing the results we can also conclude that children in this dataset interact more with peers of the same gender. Moreover, we can see that boys tend to look for interactions with older boys, whereas girls show more interactions with girls with the same age. If we focus on the *to-node* version of *playgroundB*, we can see that boys (Gender = M) are the most reached, regardless of their age. Nevertheless, the pattern with the highest score is "Gender = M ∧ Age = low". The oldest children, however, are the ones looking for more interactions according to the from-node multidigraph version. In this version, all top-3 patterns include "Age = high", despite the gender, with small differences in the scores (11.9, 10.0 and 9.7).

Since we only have two attributes in this dataset (gender and age) we generated extra features based on the networks' metrics (Section 3.2.3). The results of the comparison version of simple attributed digraph are presented in Table 9. We observe that boys tend to look for boys with a similar hub score and that girls look for girls with similar closeness. We can associate the hub score to interactions with popular children and conclude that boys prefer to interact with boys with a similar level of interactions with popular peers. Closeness, on the other hand, may imply many interactions in general, which suggests that girls prefer to interact with girls that interact with a similar amount of peers. In general, we observed that children reach peers with similar centrality measures.

4.2.5 | Implications for psychology research

From a developmental psychology perspective, our findings confirm that combining Subgroup Discovery with Network Science is an interesting and relevant approach to understanding children's behaviour during group social play. Findings of gender homophily are consistent with other studies of play and interaction in mixed-sex groups (Stehlé et al., 2013), suggesting our methods are adequate to detect this typical behaviour. Having established this, it is interesting to note that the present methods allow for detection of distinct behavioural patterns that may give insight into how individual differences in psychosocial adjustment influence social dynamics. This has been studied qualitatively or using painstaking manually coded observations that typically do not use such a fine-grained approach (Blatchford et al., 2003), (Gibson et al., 2011). This type of study can also help to identify how certain individuals (e.g., ID = 17 above) exert high levels of influence over a peer group. These types of insights may one day support clinical assessment of behavioural difficulties, or help to provide much-needed quantitative measures of the impact of environmental changes to children's play and social relations (Gibson et al., 2017).

5 | CONCLUSIONS

In this paper we proposed an approach to extract descriptive knowledge about exceptional behaviour from demographic and movement data of social interactions. We used an existing approach which combines Subgroup Discovery and Network Science techniques to find subgroups in attributed digraphs. Our main contributions include the adaptation of this approach to movement data (representing location over time) and, as such, to directed digraphs. Furthermore, we presented a methodology that combines Outlier Detection with Subgroup Discovery in order to find exceptional behaviour on individuals that may not belong to a subgroup when interacting. Accordingly, we developed a pipeline that analyses spatio-temporal data of individuals along with some of their personal and social characteristics. Then it transforms the data into attributed directed digraphs (simple and/or multidigraphs) and performs subgroup discovery. To test our approach, we applied an artificial dataset with simple characteristics to demonstrate our approach. Furthermore, we used two real-world datasets of children interacting in the school playground. The results were as expected by the experts in the domain and similar in both datasets. Nevertheless, they can add some valuable information for further social interaction analysis. Furthermore, our results indicate that the combination of the Outlier Detection measures with Subgroup Discovery provides further interesting insights, which are not enabled by standard approaches. The proposed Subgroup Discovery approach specifically enables the inclusion of spatio-temporal properties and finds interesting subgroups which provide characteristics of children that are considered outliers. This may be of utmost importance to identifying children with higher risks of unusual behaviour.

For future work, an interesting direction is given by further alternative quality measures that might be more refined to specific interaction contexts regarding the detection of particular subgroups of interactions. Furthermore, we also aim to compare the presented method with alternative approaches and to apply it to other datasets with different properties and characteristics to further explore the potentials of our approach.

ACKNOWLEDGEMENTS

This work has been partially supported by the German Research Foundation (DFG) project "MODUS" (under grant AT 88/4-1). Furthermore, the research leading to these results has received funding (JG) from ESRC grant ES/N006577/1. This work was financed by the project Kids First, project number 68639.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ENDNOTES

[1] https://shapely.readthedocs.io

[2] datamining.liacs.nl/cortana.html

DATA AVAILABILITY STATEMENT

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

ORCID

Carolina Centeio Jorge https://orcid.org/0000-0002-6937-5359

REFERENCES

Atzmueller, M. (2014). Data mining on social interaction networks. *Journal of Data Mining & Digital Humanities*, *1*.

Atzmueller, M. (2015). Subgroup discovery. *WIREs Data Mining and Knowledge Discovery*, *5*, 35–49.

Atzmueller, M. (2018). Compositional subgroup discovery on attributed social interaction networks. In *Proc. International Conference on Discovery Science* (pp. 259–275). Springer.

Atzmueller, M., Doerfel, S., & Mitzlaff, F. (2016). Description-oriented community detection using exhaustive subgroup discovery. *Information Sciences*, *329*, 965–984.

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data*. Wiley.

Berlanga, F., del Jesus, M. J., González, P., Herrera, F., & Mesonero, M. (2006). *Multiobjective evolutionary induction of subgroup discovery fuzzy rules: A case study in marketing*. Paper presented at Proc. Industrial Conference on Data Mining (pp. 337–349).

Blatchford, P., Baines, E., & Pellegrini, A. (2003). The social context of school playground games: Sex and ethnic differences, and changes over time after entry to junior school. *British Journal of Developmental Psychology*, *21*(4), 481–505.

Bondy, J. A., & Murty, U. S. R. (1976). *Graph theory with applications* (Vol. 290). Elsevier Science Ltd/North-Holland.

Breunig, M. M., Kriegel, H., Ng, R. T., & Sander, J. (2000). *LOF: Identifying density-based local outliers*. Paper presented at Proc. SIGMOD (pp. 93–104).

Cabrera-Quiros, L., Demetriou, A., Gedik, E., van der Meij, L., & Hung, H. (2018). The matchnmingle dataset: A novel multi-sensor resource for the analysis of social interactions and group dynamics in-the-wild during free-standing conversations and speed dates. *IEEE Transactions on Affective Computing*, *PP*(99), 1–17.

de Sá, C. R., Duivesteijn, W., Soares, C., & Knobbe, A. J. (2016). Exceptional preferences mining. In *Proc. International Conference on Discovery Science* (pp. 3–18). Springer.

Djenouri, Y., Belhadi, A., Lin, J. C., & Cano, A. (2019). Adapted k-nearest neighbors for detecting anomalies on spatio-temporal traffic flow. *IEEE Access*, *7*, 10015–10027.

Djenouri, Y., Belhadi, A., Lin, J. C., Djenouri, D., & Cano, A. (2019). A survey on urban traffic anomalies detection algorithms. *IEEE Access*, *7*, 12192–12205.

Duivesteijn, W., & Knobbe, A. J. (2011). *Exploiting false discoveries - statistical validation of patterns and quality measures in subgroup discovery*. Paper presented at Proc. ICDM (pp. 151–160).

Gamberger, D., & Lavrac, N. (2002). Expert-guided subgroup discovery: Methodology and application. *Journal of Artificial Intelligence Research*, *17*, 501–527.

Gibson, J., Cornell, M., & Gill, T. (2017). A systematic review of research into the impact of loose parts play on children's cognitive, social and emotional development. *School Mental Health*, *9*(4), 295–309.

Gibson, J., Hussain, J., Holsgrove, S., Adams, C., & Green, J. (2011). Quantifying peer interactions for research and clinical use: The manchester inventory for playground observation. *Research in Developmental Disabilities*, *32*(6), 2458–2466.
Goffman, E. (Ed.). (1967). *Interaction ritual: Essays in face to face behavior*. AldineTransactions.
Grosskreutz, H., & Rüping, S. (2009). *On subgroup discovery in numerical domains*. Paper presented at Proc. ECML/PKDD (p. 30).
Hawkins, D. (1980). *Identification of outliers*. Springer.
Heravi, B. M., Gibson, J. L., Hailes, S., & Skuse, D. (2018). *Playground social interaction analysis using bespoke wearable sensors for tracking and motion capture*. Paper presented at Proc. International Conference on Movement and Computing (pp. 21:1–21:8).
Herrera, F., Carmona, C. J., González, P., & del Jesús, M. J. (2011). An overview on subgroup discovery: Foundations and applications. *Knowledge and Information Systems*, *29*, 495–525.
Jones, E., Oliphant, T., & Peterson, P. (2001). *SciPy: Open source scientific tools for Python*. Retrieved from http://www.scipy.org/
Jorge, A. M., Pereira, F., & Azevedo, P. J. (2006). *Visual interactive subgroup discovery with numerical properties of interest*. Paper presented at Proc. International Conference on Discovery Science (pp. 301–305).
Jorge, C. C., Atzmueller, M., Heravi, B. M., Gibson, J. L., de Sá, C. R., & Rossetti, R. J. F. (2019). *Mining exceptional social behaviour*. Paper presented at Progress in Artificial Intelligence, Proc. EPIA Conference on Artificial Intelligence (pp. 460–472).
Kleinberg, J. M. (1999). Hubs, authorities, and communities. *ACM Computing Surveys*, *31*, 5.
Klösgen, W. (2002). *Handbook of data mining and knowledge discovery*. Oxford University Press, Inc.
Klösgen, W., & May, M. (2002). *Spatial subgroup mining integrated in an object-relational spatial database*. Paper presented at Proc. PKDD (pp. 275–286).
Konijn, R. M. (2017). *Detecting interesting differences: Data mining in health insurance data using outlier detection and subgroup discovery*. https://research.vu.nl/en/publications/detecting-interesting-differences-data-mining-in-health-insurance.
Lauw, H. W., Lim, E., Pang, H., & Tan, T. (2010). Stevent: Spatio-temporal event model for social network discovery. *ACM Transactions on Information Systems*, *28*, 15.
Lemmerich, F., Atzmueller, M., & Puppe, F. (2016). Fast exhaustive subgroup discovery with numerical target concepts. *Data Mining and Knowledge Discovery*, *30*(3), 711–762.
Lemmerich, F., & Becker, M. (2018). *pysubgroup: Easy-to-use subgroup discovery in Python*. Paper presented at Proc. ECML/PKDD (pp. 658–662).
Liu, W., Zheng, Y., Chawla, S., Yuan, J., & Xie, X. (2011). *Discovering spatio-temporal causal interactions in traffic data streams*. Paper presented at Proc. SIGKDD (pp. 1010–1018).
McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*, 415–444.
Meeng, M., & Knobbe, A. (2011). *Flexible enrichment with cortana–software demo*. Paper presented at Proc. BeneLearn (pp. 117–119).
Messinger, D. S., Prince, E. B., Zheng, M., Martin, K., Mitsven, S. G., Huang, S., Stölzel, T., Johnson, N., Rudolph, U., Perry, L. K., Laursen, B., & Song, C. (2019). Continuous measurement of dynamic classroom social interactions. *International Journal of Behavioral Development*, *43*, 263–270.
Newman, M. (2010). Introduction: A short introduction to networks and why we study them. In *Networks: An introduction*. O.U.P.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., VanderPlas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.
Qu, J. (2008). *Outlier detection based on voronoi diagram*. Paper presented at Proc. International Conference on Advanced Data Mining and Applications (pp. 516–523).
Roddick, J. F., & Spiliopoulou, M. (1999). A bibliography of temporal, spatial and spatio-temporal data mining research. *ACM SIGKDD Explorations Newsletter*, *1*, 34–38.
Romero, C., González, P., Ventura, S., del Jesús, M. J., & Herrera, F. (2009). Evolutionary algorithms for subgroup discovery in e-learning: A practical application using moodle data. *Expert Systems with Applications*, *36*, 1632–1644.
Shekhar, S., Lu, C., & Zhang, P. (2003). A unified approach to detecting spatial outliers. *GeoInformatica*, *7*, 139–166.
Shi, Y., Gong, J., Deng, M., Yang, X., & Xu, F. (2018). A graph-based approach for detecting spatial cross-outliers from two types of spatial point events. *Computers, Environment and Urban Systems*, *72*, 88–103.
Skrlj, B., Kralj, J., Vavpetic, A., & Lavrac, N. (2017). *Community-based semantic subgroup discovery*. Paper presented at Proc. NFMCP@PKDD/ECML (pp. 182–196).
Stehlé, J., Charbonnier, F., Picard, T., Cattuto, C., & Barrat, A. (2013). Gender homophily from spatial behavior in a primary school: A sociometric study. *Social Networks*, *35*(4), 604–613.
Terry, M. A., Mynatt, E. D., Ryall, K., & Leigh, D. (2002). Social net: Using patterns of physical proximity over time to infer shared interests. In *CHI extended abstracts* (pp. 816–817).
Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.
Wrobel, S. (1997). *An algorithm for multi-relational discovery of subgroups*. Paper presented at Proc. PKDD (pp. 78–87). Berlin/Heidelberg, Germany: Springer.
Zwilling, C. E., & Wang, M. Y. (2014). *Multivariate voronoi outlier detection for time series*. Paper presented at Proc. IEEE Healthcare Innovation Conf. (pp. 300–303).

AUTHOR BIOGRAPHIES

Carolina Centeio Jorge received her Master's in Informatics and Computing Engineering from the Faculty of Engineering of University of Porto (FEUP) in 2019. On the side, she also took part in research projects both at University of Porto and INESCTEC. Her Master thesis was done in collaboration with University of Twente and focused on Exceptional Behaviour Discovery, where she used data mining techniques to study patterns in social interactions. After graduating, Carolina took part in Vulcanus in Japan program, where she could experience a traineeship in Computer Vision and Deep Learning at Omron Sinic X Corp., in Tokyo, Japan. Since October 2020, Carolina is a PhD student at the

Interactive Intelligence group of TU Delft. Her PhD has a special focus on mental models in the context of human-AI teams, such as enabling artificial agents (e.g., robots) to understand and predict human teammates and effectively act accordingly. Carolina's main research interests lie in Behaviour Analysis and Pattern Recognition.

Martin Atzmueller is Full Professor (W3, tenured) at the Institute of Computer Science at Osnabrück University (Germany), where he holds the ROSEN-Group-Endowed Chair of Semantic Information Systems and heads the Semantic Information Systems research group. Previously, he also held appointments at Tilburg University (The Netherlands) as an Associate Professor, and at the Université Sorbonne Paris Nord (France) as a Visiting Professor; also, he earned his habilitation (Dr. habil.) in 2013 at the University of Kassel, where he also was appointed as adjunct professor (Privatdozent). Martin Atzmueller's research interests include Artificial Intelligence, Complex Networks, Knowledge Discovery, Machine Learning and Human-Centered Data Science. His work focuses on how to 'make sense' of complex information and knowledge processes - leveraging the massive amounts of data collected in science and industry by intelligent analytics and semantic interpretation. By connecting computational approaches with the human cognitive, behavioral, and social contextual perspectives - thus linking technologies with their users - the goal is to augment human intelligence and to assist human actors in all their purposes, both online and in the physical world.

Behzad M. Heravi is Director of Technology and Innovation at Vortex IoT. He leads product R&D and both the sourcing and implementation of Vortex's innovation led technology strategy. Behzad is an award-winning Computer Scientist highly skilled in product R&D, with a PhD in Communication Systems, MSc in Aerospace Engineering and a BSc in Mechanical Engineering. Behzad was previously a research fellow at UCL and is both passionate and driven by innovation and embracing emerging technologies. He is a published research author in Artificial Intelligence, Machine Learning, Statistical Inference and Signal Processing. Behzad has a long and accomplished history in R&D and has successfully secured several funding awards from Innovate UK.

Jenny L. Gibson is Senior Lecturer/Associate Professor in Psychology and Education at the University of Cambridge. She is PI of the Play and Communication Lab and part of the leadership team for the centre for research on Play in Education, Development and Learning (PEDAL). Jenny completed her doctoral research at the University of Manchester and subsequently held a postdoctoral post at the University of Cambridge department of psychiatry. Jenny has a long-standing interest in understanding how children's playful behaviour relates to other aspects of their development. Her research focus is on the social dynamics of playful interactions and how these might differ in neurodivergent children compared to children following more typical developmental trajectories. Jenny is interested in how computational approaches can connect with those from humanities and social sciences to further our understanding of the complexities of human behaviour.

Rosaldo J. F. Rossetti is a senior research fellow and member of the directive board of the Artificial Intelligence and Computer Science Lab and a faculty member of the executive committee of the Department of Informatics Engineering at the University of Porto, Portugal, where he is an associate professor. Dr Rossetti served as an elected member of the Board of Governors of the IEEE ITS Society during term 2011-2013 and was a member of the steering committee of the IEEE Smart Cities Initiative from 2013 to 2017. He is currently chair of the IEEE ITSS Artificial Transportation Systems and Simulation Technical Activities Committee, for which he was the recipient of the Best TAC of the IEEE ITS Society Award in 2017. He is an associate editor of the IEEE Transactions on Intelligent Transportation Systems and the ITS Department editor of the IEEE Intelligent System Magazine. He is also a member of ACM, APPIA (the Portuguese AI Society), and the European Social Simulation Association. Dr Rossetti's main research interests include behavioural modelling, social simulation, and spatiotemporal data analytics and machine learning. He focuses on applications of multiagent systems as a modelling metaphor to address issues in artificial transportation systems and simulation, future mobility paradigms and urban smartification, and explores the potential uses of serious games and gamification in transportation and mobility systems. He holds a PhD in Computer Science from INF-UFRGS, Brazil (2002), having carried out his doctoral work as a research student at Leeds University's Institute for Transport Studies, UK.

Cláudio Rebelo de Sá is a senior data scientist working at the University of Twente. I have a PhD in Computer Science from the University of Leiden. My main areas of expertise are Data Mining, Neural Network, Deep Learning and Computer Vision.

How to cite this article: Centeio Jorge C, Atzmueller M, Heravi BM, Gibson JL, Rossetti RJF, Rebelo de Sá C. "Want to come play with me?" Outlier subgroup discovery on spatio-temporal interactions. *Expert Systems*. 2021;1–20. https://doi.org/10.1111/exsy.12686