The Visual Analytics Approach for Analyzing Trajectories of Critical Infrastructure Employers †

Evgenia Novikova 1,2, Igor Kotenko 2,* and Ivan Murenin 2

1 Department of Information Systems, Saint Petersburg State Electrotechnical University, 197022 Saint Petersburg, Russia; novikova.evgenia123@gmail.com
2 Laboratory of Computer Security Problems, Saint Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, 199178 Saint Petersburg, Russia; imurenin@gmail.com
* Correspondence: ivkote@comsec.spb.ru
† This paper is an extended version of Visualization-Driven Approach to Anomaly Detection in the Movement of Critical Infrastructure; Lecture Notes in Computer Science, 10446, Springer, Cham, 50–61, 2017. doi:10.1007/978-3-319-65127-9_5; and Visualizing anomalous activity in the movement of critical infrastructure employees; 2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), St. Petersburg, Russia, 1–3 February 2017. (https://ieeexplore.ieee.org/document/7910602).

Received: 27 June 2020; Accepted: 18 July 2020; Published: 1 August 2020

Abstract: Employees of different critical infrastructures, including energy systems, are considered to be a security resource, and understanding their behavior patterns may leverage user and entity behavior analytics and improve organization capabilities in information threat detection such as insider threat and targeted attacks. Such behavior patterns are particularly critical for power stations and other energy companies. The paper presents a visual analytics approach to the exploratory analysis of the employees’ routes extracted from the logs of the access control system. Key elements of the approach are interactive self-organizing Kohonen maps used to detect groups of employees with similar movement trajectories, and heat maps highlighting possible anomalies in their movement. The spatiotemporal patterns of the routes are presented using a Gantt chart-based visualization model named BandView. The paper also discusses the results of efficiency assessment of the proposed analysis and visualization models. The assessment procedure was implemented using artificially generated and real-world data. It is demonstrated that the suggested approach may significantly increase the efficiency of the exploratory analysis especially under the condition when no prior information on existing employees’ moving routine is available.

Keywords: visual analytics; data mining; moving entities; route patterns; anomaly detection; self-organizing Kohonen maps; Gantt chart-based visualization; heat maps

1. Introduction

Application of digital technologies has greatly contributed to economical and social aspects of humanity development and enhanced environment protection, however, it has been shown that socio-technical systems and ecosystem has become more vulnerable to cyber-threats, and especially this problem is particularly important to critical infrastructure in the energy sector [1] as any system malfunctioning may cause severe impact on all aspects of human living. The resilience of the any energy company to cyber security threats is a critical issue in provision of its safe and sustainable functioning. Insider threat is one of the sophisticated information security threats requiring the organization to establish a process for tracking unusual behavior of employees or potential incidents [2]. Analysis of employees’ movement allows one to discover existing safety, access control policies, business needs and operational characteristics of energy companies which may impact critical infrastructure resilience. Therefore, understanding these patterns can significantly contribute to the resilience enhancement and detection of possible threats.
routines and supports monitoring their compliance to the prescribed ones [3,4]. Understanding of the employees’ behavior patterns may significantly enhance user and entity behavior analytics and improve organization capabilities in information threat detection such as insider threat, financial frauds and targeted attacks [5,6].

Analytical techniques implemented in modern access control systems and employee activity monitoring systems are mostly targeted for employee productivity assessment by calculating working hours, estimating dynamics of employee lateness and evaluating user activity in work places [7–9]. At the same time, according to [10], the analysis of user activity should allow one to construct the behavior patterns, supporting identification of various anomalies that may be the signs of the sophisticated security threats such as insider threats and targeted attacks.

The paper presents an approach to constructing the spatiotemporal patterns of the employees’ movement inside the organization and detection of anomalous deviations in their routes using access control system logs. A set of interactive visualization models tightly coupled with data mining techniques assists in establishing groups of employees with similar trajectories, spatiotemporal patterns of the employees’ routes and detecting anomalous deviations in their movement. Specifically, the authors’ contribution is a visual analytics approach to the analysis of the movements of critical infrastructure stuff that operates heterogeneous data, including employee position, location of their work places and temporal attributes of the movement to support tasks of the cyber–physical security. The key elements are self-organizing Kohonen maps (SOMs) used to detect groups of employees with similar behavior and days when they move similarly. The SOMs are supplemented by a specially designed glyph that allows discovering periodicity in their trajectories depending on a day of week, and anomaly heat map equipped with an anomaly ranking mechanism.

This paper presents the development of the results partly presented on the MMM-ACNS 2017 [11] and ElConRus conferences [12] and in the article [13]. It contains a formal description of the analysis process, including an extended description of the developed visual analytics techniques and rationale for their choice, recommendations to selection of the analysis models parameters and presents an expert assessment on the effectiveness of the proposed visualization techniques as well as the efficiency of the approach in the anomaly detection.

The rest of the paper is structured as follows: Section 2 presents state-of-the-art in visual analysis of the trajectories and discusses implemented techniques in the access control and employee activity monitoring systems. Section 3 describes the approach suggested, including general scheme of the analytical process, visualization models and interactions with them. Section 4 outlines the case studies used to assess the proposed approach and discusses the obtained results. Section 5 considers and discusses the results of evaluating the efficiency of the proposed visual analytics technique. Section 6 sums up the authors’ contributions and defines directions of the future work.

2. Approaches to Anomaly Detection in Objects Movement (Related Work)

One of the tasks of the trajectory mining is discovering and specifying the movement patterns of the moving objects existing in trajectories. It gives information about when and where the pattern occurs, and what entities are involved. Clustering is a common approach to the extraction of the movement patterns involving groups of objects moving together and detection of possible interactions between group members especially when there is no possibility to obtain patterns of normal behavior by making observations on normal and abnormal behavior of the entities [14,15]. The obtained clusters are then used to describe the normal behavior model for anomaly detection. There is a variety of clustering algorithms based on similarity, distance, density or distribution assessment [16,17]. These techniques may discover interesting behavioral patterns and anomalies but in the most cases they need to be supported by the visualization techniques explaining the final result.

The self-organizing map (SOM) is an artificial neural network that implements multidimensional data clustering and produces low-dimensional projection of the input space preserving distance between samples. Shreck et al. [18] applied SOM to analyze object trajectories in some abstract space.
and propose a visualization-driven framework that supports a user-guided SOM initialization process. In Ref. [19], authors proposed an approach for analysis of spatially and temporally referenced attribute values. It supports two complementary views on spatial-temporal data called “spatial situations” that focus on spatial aspects of data and “local temporal variations” that focus on temporal aspects of data referenced to some location. SOM is used to solve both of these tasks. To simplify the analysis of the SOM output, authors supplemented each SOM view with a special matrix-based image that provides information either on temporal or spatial attributes of an object. In Ref. [15], a modified SOM, named as Density-based Simultaneous Two-Level Self Organizing Map, is applied to reveal the social organization of the ant colony.

It has been shown that the analysis of multidimensional data requires the development of the visual analytics frameworks implementing various visualization models that propose different points of view on the source data and, thus, form the comprehensive understanding of data being analyzed [19,20].

In the general case, there are three main approaches to graphical representation of moving object trajectories:

1. Interactive maps, often supplemented with glyphs encoding movement or object attributes;
2. 3-dimensional space-time cubes;
3. Stacking-based visualizations [21].

The most natural way to represent the location aware data is to use geographical maps and plans. The routes in this case are shown as lines, and their attributes or object characteristics are encoded either by color or specially designed glyphs [22–26]. For example, an interesting approach to display the high dimensional spatial attributes and statistics associated with different routes is presented in [25]. The authors place map visualization of the vehicle routes for a given source/destination pair in the center of the circular area and surround it by the histogram that shows statistics on a number of trajectories and their attributes recorded during a particular time of day, thus reflecting the existing dependence between trajectories of the vehicles and time.

Kruger et al. added semantic information to the movement data extracted from microblog services and social location service [27]. These data are encoded by pictograms or word clouds depending on type of data.

The flow maps are used when the exact trajectory is not important [23]. They focus analyst attention on determining destinations and sources of the routes. The numerical attributes of the flows are usually encoded by two visual variables—line width and color. Though map-based visualizations intuitively convey spatial attributes of the movement, the representation of the temporal attributes of the trajectories are rather difficult. In Ref. [28], authors proposed a visual analytics approach to investigate how massive movement flows change over time, it is based on a graph-based technique that includes a spatial aggregation step with a consequent temporal clustering step to reduce input graph volume.

3D visualization models such as space-time cube allow displaying spatial and temporal characteristics of the movement simultaneously. The trajectory is represented as a line in the three dimensional space, where the vertical axis usually stands for time, and the horizontal plane is used to map spatial attributes. In Ref. [29], an interesting 3D visualization technique to explore a set of trajectories is presented. The authors used two dimensions to represent geographical data, but instead of using third dimension as a time axis, they reserved a portion of it for each trajectory. Thus, all trajectories are displayed sequentially as a stacked band. The numerical attributes of trajectories values are encoded by color. Like all 3D visualizations, the space-time cubes could be ineffective because of occlusions and cluttering of the trajectories.

Stacking-based visualizations of the routes are based on time line. One axis represents the time, and another axis represents attributes of the moving object or its trajectory, then the route is represented by a curve or polyline. To display a set of trajectories, stacked curves or polylines synchronized in time are used [30]. The Gantt chart-based trajectories visualization techniques can be considered as
a type of the stacking-based visualization techniques when its vertical axis is used to list different moving objects or visited locations, while the horizontal axis denotes time. The corresponding bars are usually divided into segments colored according to values of trajectory attributes. In Ref. [31], authors discussed the applicability of the Gantt chart-based visualization techniques to reveal possible contacts between moving objects and showed that it is useful in the analysis of the interactions between small groups of entities and can reveal abnormal behavior of the entity.

In Ref. [32], the Gantt-chart based visualization named Event Quiltmap is used to analyze movement of the employees within organization. Unlike to traditional Gantt chart the vertical axis of the suggested visual model represents time of the day, while the horizontal axis denotes a set of days being analyzed. Thus, each band segment stretches vertically for the duration of time under which the employee did not trigger an event of the proximity switches. The cells use color to encode the zone number consistently across the building floors, and texture to encode the building floor level, which gives the map the appearance of a quilt. The Event Quiltmap visualization is supported by a 3D Building View which represents itself a 3D plan of building and shows employee trajectories throughout a day.

Analysis of the employees’ movement in the modern access control system includes constructing the employees’ routes within the building and movement heat maps reflecting the most visited controlled zones. The main focus is done on calculating of distances walked or driven by vehicle, and monitoring the compliance with temporary movement regulations describing schedules of visiting of the controlled zones [8,9]. Some systems allow detection of the anomalies concerning with cases of one-time registration of several co-workers by one proximity switch or triggering proximity switch events for an absent employee [8]. In the general case, existing access control systems do not form employees’ movement patterns based on analysis of their trajectories. The anomaly detection is done using rules that are constructed according to the employees’ role profile. In addition, the analysis of the security incidents consists in working with the raw proximity switch logs, presented in a tabular format or in video format, no visual analytics techniques are introduced to support visualization driven interactive analysis of the data.

The presented approach in this paper on the analysis of the employees’ trajectories is close to the approach presented in [32]. The visualization technique named BandView proposed in the approach to investigate the raw data about employees movement is quite similar to the Event Quiltmap described in [32] as they both are based on Gantt chart. However, unlike [32], the authors use SOM clustering technique for revealing possible patterns in employees’ movement and apply statistics-based mechanism for ranking detected deviations in employee’s route considering the periodicity of their occurrence. Unlike [15], the SOM is applied to the analysis of trajectories extracted from the proximity switch logs, and considers both spatial and temporal attributes of the movement. The graphical presentation of the SOM is enforced by the special glyph that gives brief characteristics on entities belonging to one cluster.

3. An Approach to Analysis of the Employees’ Movement

The goal of the proposed approach to the exploratory analysis of the employees’ movement is to extract and analyze the information on common behavior of the organization employees expressed in terms of their routes within building, visited zones and interactions between co-workers. Initially, this research was motivated by the 2016 VAST Mini Challenge 2, however, the proposed solution is applicable to the wide variety of analytical tasks concerning staff movement monitoring. The analysis of the employee’s movement patterns allows an employer to develop better workspace layout, enhance energy usage by tuning heating, ventilation and conditioning system lights to the scheme usage of a certain areas [33]. The primary goal of the suggested approach is to enhance the organization resilience to insider threat by providing insight on how people move within the organization and revealing suspicious deviations in their routes. For example, according to [2], employees are viewed as security resources and it is recommended that critical infrastructure organize a process for monitoring
unusual behavior. That is why when designing this approach the authors surveyed security specialists to understand what questions they may be interested in when analyzing employees’ movement. The conducted survey allowed us to establish the following questions:

1. Are there any groups of employees sharing similar behavior? Does it depend on the employees’ position in the organization?
2. What is the typical daily route of the co-workers belonging to one group? Does it change depending on the week day?
3. Are there any specific interactions between co-workers belonging to one or different department?
4. Are there any deviations in employee’s route?
5. What is the character of the detected deviations, i.e., how often, when and where did they take place? Who else was involved in a particular anomaly?

These questions define an employee-centric character of the analysis process and its main steps. These steps are as follows:

1. Data preprocessing step, preparing entries of the proximity switch logs for further analysis;
2. Detection of the groups of the employees with similar behavior and their visualization using SOM. Validation of the clustering results by viewing raw data presented by a specially designed Gantt-based visualization model named BandView;
3. Detection of the periodicity in the movement of the employees belonging to one group by revealing week days with similar routes. Graphical representation of the detected groups using SOM;
4. Detection of the anomalies in the employee’s routes by the statistical assessment of the trajectory deviations from patterns and graphical representation of the anomalies using heat map;
5. Detailed analysis of the anomalous part of the employee route using the BandView visualization model.

The analysis scheme given in context of information seeking mantra—“overview first—zoom and filter—details on demand” [34] is shown in Figure 1. The analysis steps and supporting visualization and interaction techniques are discussed in detail in the subsections below.

3.1. Data Preprocessing Step

The source data in the presented approach to the analysis of the employees’ movement within the organization are the proximity switch logs that contain data about the time of visiting the monitored zone by a certain employee. The important requirement to the source logs is the possibility to extract from them the following information: <employee_id, controlled_zone_id, timestamp>. Thus, it is not important what kind of technology is used to implement the access control system. Logs could be obtained from the rfid/nfc proximity card readers [35] or an application interacting with Bluetooth beacons [8]. Information on the employees’ department or position as well as the plan of the controlled zones allow more accurate interpretation of movement patterns by adding a semantic component to them, for example, by indicating that the workplace of the employee is located in the given zone.

The distinctive feature of the logs generated by the proximity card readers is that they appear irregularly when an employee enters or leaves the controlled zone, as the result the interval between log entries may vary from tens of a second to tens of hours. Moreover, some employees move a lot during their working day due to their job responsibilities, while others mostly stay at their work places. Thus, the proximity switch logs of sensors can be considered as time-series with variable length and gap between values. Figure 2 shows logs of the proximity switches as time series, the X axis corresponds to time, while the Y axis denotes to an identifier of the controlled zone, it shows routes of three employees, belonging to one department.
The activity of the employee during selected time interval is assessed by analyzing the number of visits of a particular controlled zone and the duration of staying in it. The representation of the proximity switch logs as time series.

Figure 1. Scheme of the proposed visual analysis process.

Figure 2. The representation of the proximity switch logs as time series.

One of the methods of such time-series analysis is segmentation when an input time-series is split into a sequence of discrete segments and calculating average parameters characterizing these segments in order to reveal its underlying properties [36]. In this research the authors adopt this approach to evaluate spatial and temporal attributes of a route during each discrete time period. The activity of the employee during selected time interval is assessed by analyzing the number of visits of a particular controlled zone and the duration of staying in it.

Let \( E = \{e_i\}_{i=1}^{n} \) be a set of employees, \( Z = \{z_j\}_{j=1}^{m} \) a set of the controlled zones, \( T = \{t_k : t_i < t_j ; i < j\}_{k=1}^{n} \) an ordered set of the timestamps, then a tuple \( (e_i, z_j, t_k) \) denotes an entry...
of the proximity switches log. The set of log entries can be then defined as \( \text{LOGS} = \{(e_i, z_j, t_k)\}, \)
\(i = 1 + n, j = 1 + m, k = 1 + p.\) Let \(T_0\) denote the time interval, presented by the first and last entries of the log. The time interval \(T_0\) is split into the sequence of the equal time slots \(\Delta t:\)
\(T_0 = \{\Delta t_i : \Delta t_i = \Delta t_{i+1} = [t_i, t_{i+1}) \}; i \neq j; i, j \leq l; \}_{i=1}^r.\)

For each time slot \(\Delta t_i\) and each employee \(e_i\), a pair \((n_{z_j}^\Delta t_i, \Delta t_j^\Delta_i)\) is calculated, where \(n_{z_j}^\Delta t_i\) denotes a number of visits and \(\Delta t_j^\Delta_i\) denotes a duration of staying in \(z_j\) controlled zone. Thus, the set \(\text{LOGS}\) is transformed to a set of ordered pairs \(\text{LOGS} = \{(n_{z_j}^\Delta t_i, \Delta t_j^\Delta_i)\}_{i=1}^r, j = 1 + m, l = 1 + r,\) calculated for each employee \(e_i, i = 1 + n.\) This set of the ordered pairs forms a set of spatiotemporal attributes of the employee’s trajectory. The attributes are ordered by time slots firstly and then by controlled zones.

The \(\Delta t\) parameter controls the granularity of the analysis. By setting it equal to the \(T_0\) interval, the analyst gets the summarized information on how the employee visits zones and average time spent inside them. The authors recommend setting it to a more meaningful time interval such as month, week, day or hour depending on duration of time period described by available logs. This allows revealing how attributes describing employees’ movement change over the time, and assessing the character of these changes. For example, setting \(\Delta t\) equal to day or week, it is possible to detect days or weeks with similar routes of the employee. If the revealed similarity depends on week day or week (odd or even) it is possible to conclude that there is a periodicity in the employees’ movement depending on week day or week day of week. Setting value of \(\Delta t\) less than a day, for example, 4, 8 or 12 h makes it possible to discover differences in movement depending on time of the day. These differences may occur when employees work in several shifts.

Thus, to obtain meaningful analysis results it is recommended setting \(\Delta t\) equal to divisors of 24 if time unit is hour, or equal to expected duration of periodicity in movement, for example, 1 or 2 weeks, 1 month, etc. It also should be less or equal to duration of time presented by the logs.

This data preprocessing step allows transforming initial set of proximity switch logs to a set of numeric vectors of a finite length and applying different clustering as well as visualization techniques for extracting groups of employees with similar behavior.

### 3.2. The SOM-Based Views

When choosing clustering and visualization techniques for extracting group of employees with similarities the authors tried to take into account following considerations: (1) the number of clusters is unknown; (2) the results of clustering should be easy to interpret; (3) they should include information on existing difference, i.e., distances, between clusters as well as clusters capacity; (4) it should be possible to extract cluster’s centroids to evaluate deviations of the cluster’s members from it.

The visual clustering techniques do not require a prior knowledge about number of clusters. There are different approaches based on axis-aligned projection, linear projection and manifold learning for implementing visual clustering [37]. However, these techniques do not reduce the size of the original dataset. Therefore, they are good as a starting point in relation seeking tasks and for the validation of the automated analysis models’ output. They do not produce any average description of the cluster such as cluster centroid and, therefore, could not be used for pattern extraction from trajectories. The majority of centroid-based clustering techniques require a number of clusters as input parameter of the algorithms.

To detect groups of employees with similar behavior and outliers, the authors suggest using SOMs known also as Kohonen maps [38]. It is an artificial neural network that is trained using unsupervised learning to map multidimensional input data into a low-dimensional (typically two-dimensional) space. The SOM consists of nodes or neurons associated with weight vectors. The nodes are usually arranged in a hexagonal or rectangular grid. The weight vectors have the same dimension as input data space. During the iterative learning, the input vectors are compared to the weight vector of each neuron, and the weights of the neuron best matching to the input vector and weights of the nearby neurons are adjusted to be closer to the input vector. This makes the SOM topology-preserving map,
and the data samples assigned to the adjacent nodes of SOM are more alike than data samples assigned to nodes that are far from each other. Training SOM requires data without missing values for each attribute, but the suggested data preprocessing step produces vectors with values for each dimension.

The most commonly used visualization technique to present SOMs is the U-Matrix [39]. It shows a data structure by displaying the average distances between weight vectors of neighboring units. In the approach authors use a hexagon grid and, therefore, the vector of 6 adjacent nodes is considered. The darker color of the node, the more it differs from the neighbors. The adjacent light nodes are quite similar to each other. Nodes containing clusters’ centers are marked with the circle glyph. Its size reflects the number of objects in the cluster.

In the approach the SOMs are used twice—to detect groups of employees with similar routes (Employees SOM) and to reveal periods of time when employees move alike (Periodicity SOM). The attribute vector describes activity of the employee during all periods of time being analyzed, and deviations in the movement taking place once or rarely do not influence on the result of the clustering. This allows authors to assume that this SOM reflects differences in movement existing due to peculiarities of the employee’s job responsibilities in the organization.

The granularity of the analysis depends on the number of the SOM nodes—the bigger the SOM size, the more data clusters are displayed. In this case employees having relatively similar behavior would be located in the region of the SOM marked by similar light grey color. The experiments showed that the differences in the employees’ behavior belonging to different but adjacent nodes are explained in most cases by the variance in the duration of staying within controlled zones, wherein the sequence of visits of the controlled zones is similar. The variance of the visit durations depends on the duration of the time interval chosen on the preprocessing step. It is recommended to select the size of the SOM comparable to the number of employees being analyzed. However, an analyst has a possibility to change the dimensions of the SOM to see how the employees are grouped.

The SOM presented in Figure 3 shows the results of clustering of employees belonging to one department, circles mark the SOM nodes with clusters. The size of circle reflects the number of employees in the cluster. It is clearly seen that there are seven groups of co-workers having the similar movement pattern. One of these groups is rather numerous (group 2), while others consist of one-two persons. The groups 1–3 have rather similar trajectories as the colors of their cells are almost alike, the employees belonging to the groups 4, 5 have slight differences in the routes as the colors of the corresponding SOM cells are slightly darker then cells with the first three clusters. Two groups of employees 6, 7 displayed in the lower left corner of the SOM expose significantly different behavior as they are separated from the other by a set of dark nodes. Further analysis of the movement patterns of these employees showed that the main differences in their behavior is explained by the time they start their work day: their working day starts in the second part of the day, while others employees start working in the morning.

The conducted survey showed that the security analysts are interested in detection of periodicity in the employees’ movement. This periodicity may arise due to the fact that employees have responsibilities implemented on some regular basis depending on the day of week. The primary goal of the second SOM—Periodicity SOM—is to reveal time intervals with similar behavior for the selected group of employees allowing thus an analyst to reveal periodicity in their movement. To construct the Periodicity SOM view, the weight vectors of the cluster centroids produced by the Employees SOM are transformed to a set of vectors corresponding to the days by splitting source vectors into a set of vectors of smaller length. The length of the resulting vector depends on the time interval \( \Delta t \) set at the data preprocessing step. If \( \Delta t \) is set equal to divisors of 24 and time unit is hour, then the length \( L \) of the resulting vector is defined as \( L = n_z \times 2 \times (24 / \Delta t) \), where \( n_z \) is the number of the controlled zones. Multiplier of 2 is used because two parameters—number of visits and duration of staying—are calculated for each controlled zone. However, the approach allows setting \( \Delta t \) in another time units such as week or month. In this case the length \( L \) of vectors is defined as follows \( L = n_z \times 2 \), and the results of clustering could not be used to reveal week periodicity in the employees’ movement. If any
periodicity is detected in this case, then it depends on the initial value of $\Delta t$ and may be hard to be interpreted. Thus, the authors recommend using the Periodicity SOM only when the time interval $\Delta t$ is set to divisors of 24 and time unit is hour. This allows an analyst to determine what days of the week are contained in each group of days and thus discover dependency of the employees’ route from the day of the week.

![Figure 3. The Employees self-organizing Kohonen map (SOM) view.](image)

The result of clustering is also displayed using U-Matrix. The SOM nodes are optionally may be complemented with a special glyph named WeekCircle that displays the distribution of the days in the cluster according to the day of the week. The authors designed it to highlight movement patterns having periods equal to one or two weeks. The 14-day period is considered as in many organizations the employee’s activity routine depends on the type of the week in the year—odd or even. Depending on the mode—7 days or 14 days—the glyph is divided into 7 or 14 sectors correspondingly. The scheme of the glyph is presented in Figure 4. It should be noted that this option is available in the case when $\Delta t$ is set to divisors of 24 and the time unit is hour.

![Figure 4. WeekCircle glyph represented in two modes: (a) and (b) glyphs are depicted in 7-day mode and (c) and (d) glyphs are depicted in 14-day mode.](image)

The two leftmost glyphs are implemented in 7-day mode, the glyph in Figure 4a shows that the cluster contains routes taken place every Monday, Wednesday and Friday. The glyph in Figure 4b shows that the cluster consists of the routes taken place on the weekends. The two rightmost glyphs are implemented in the 14-day mode. The right half of the WeekCircle represents the odd week, and the left half of it represents the even week. The Mondays are displayed by the top sectors and Saturdays by the bottom sectors, thus the odd week is in mirror reflection with the even one. Like the glyph in Figure 4b, the glyph in Figure 4d indicates that the group of days is represented by the weekends only. The glyph in Figure 4c shows that the cluster consists of Mondays, Wednesdays and Fridays of the odd week, meaning that the employee or group of employees has particular duties implemented every second Monday, Wednesday and Friday. Thus, the usage of the WeekCircle glyph may reveal some additional information on existing periodical patterns. It should be noted that the Periodicity SOM View allows also detecting days with anomalous behavior if these anomalies have long term
character, i.e., their duration is comparable to the duration of $\Delta t$. The day with such type of anomaly would constitute a separate cluster located in the close neighborhood to another cluster. The clusters consisting of only one element on the Periodicity SOM are highlighted by red border of the glyph.

### 3.3. The BandView Visualization Model

The goal of the BandView visualization model is to present raw data about employees’ movement or patterns that are extracted from their trajectories. It is a Gantt chart-based visualization model. The horizontal axis of the diagram corresponds to the time, the vertical axis shows the employees. The employees’ trajectory is represented using a segmented band. Each band segment indicates the presence of the employee in the controlled zone, and the segment length depends on the duration of staying in it. Thus, the BandView model displays both spatial and temporal attributes of the employees’ trajectory. The color of the segment is used to encode the zone’s attributes.

For example, to encode the location of the zones inside the building the following color scheme is proposed. Each floor of the organization building is assigned a certain color, the selected colors constitute qualitative color scheme best suited to representing nominal and categorical data [40]. The palette for the zones located on one floor or belonging to one organization location is created by changing saturation of the “floor” color, forming thus a sequential scheme that is more appropriate for the ordered data that progress from low to high. The greater number (ID) of the controlled zone, the darker the color of the corresponding segment. Such approach to color coding of the controlled zones enables to detect where the employee spends most of the time, how diverse their route is. Figure 5 shows the routes of the co-workers belonging to one department for one day. The zones of the first floor are displayed in brownish colors, zones of the second floor—in olive colors, and zones of the third floor—in blue colors. It is clearly seen, that almost all employees except two of them spend their work time on the second floor, two co-workers visit the third floor. Five employees start their work day in the second half of the day. Most of them spend their lunch time at the first floor, and only two of them remain on the second floor. The labels with cluster numbers are given to show how the SOM grouped the given routes of the employees. The SOM map itself is shown in Figure 3.

![Figure 5. The graphical representation of the employees’ routes using the BandView model.](image-url)

Obviously, the duration of staying in the zones varies greatly—it can be equal to tens of seconds as well as to tens of hours. This results that the lengths of BandView segments can be very small, almost invisible; while others are very long requiring much scrolling. In order to provide an analyst with the ability to view both short and long segments, a zooming mechanism based on nonlinear time scale transformation is proposed. It increases short time intervals and shortens the long ones.

To support investigation of the raw data the authors implemented filtering mechanism that allows constructing filters as logical expressions that use available movement attributes, e.g., employee ID, department, office, zone ID, location, etc. Each segment of the model is clickable; therefore, it can be expanded or collapsed. If time interval is set to a divisor of 24 and time unit is hour then BandView model is complemented by a WeekCircle model.
department, office, zone ID, location, etc. Each segment of the model is clickable; detailed information about it such as duration, timestamp, zone ID, its location, etc. is displayed in the tabular view. Thus, BandView model could be used to describe the character of the anomaly detected—where, when it took place and how long it lasted. BandView model is also used to present route patterns, in this case the Y-axis corresponds to the time intervals characterized by similar routes. If time interval is set to a divisor of 24 and time unit is hour then BandView model is complemented by a WeekCircle model.

3.4. The Graph Visualization Model

Maps or floor plans naturally convey information on existing spatial patterns of trajectories of the moving objects. However, when presenting routes of the employees using 2D plan of the multi-story building, a problem may arise concerning depiction of employee passages from one floor to another. The 3D visualization of the building may solve this problem, but as all 3D visualization techniques, the 3D building plan needs implementation of the interaction techniques such as 2D-projections, i.e., construction of the 2D floor plans that have the same problem. The authors propose to display visited controlled zones using a node-link diagram. The controlled zones are represented by graph nodes, and edges link adjacent zones. Such graph can be constructed on the basis of proximity switch logs or a map of controlled zones if the latter is unavailable. Analysis of the zone map can be useful as it allows assigning different attributes to the zones, such as presence of the café, elevators, stairs, meeting rooms, etc. These features can be used to explain the motif of the employee movement.

The zones visited by the employee are colored according to the color scheme used in the BandView visualization model, unvisited zones are left in light grey.

Apart from displaying information on the zones visited by the employee, the graph is used to convey statistical information about employees’ movements—mean duration of staying within controlled zone per day, its variance, average number of visits per day and overall number of visits during all time period being analyzed. This information is also displayed in the tabular view by clicking on the graph vertexes. It is partly written on the graph vertexes (mean duration of staying and number of visits per day). The vertex size is defined either by a number of visits or mean duration of staying in the zone. The authors consider this option helps to spot zones where the work place of the employee is located and zones visited rarely and easily. Figure 6 shows the graph of visited zones which vertexes’ size depends on number of visits during all periods of time. It is clearly seen that the employee spends most of his/her working hours in the zone 2-1—377 min or 6.27 h approximately. The analysis of the zone’s features pointed out that it contains offices thus it is possible to conclude that working place of the employee is located there. It is also clearly seen that he or she regularly visits zone 2-6 spending there almost two hours per day (110 min). This zone contains both offices and meeting room thus the authors may conclude that regular meetings are held there. Zone 1-2 is another zone with rather long mean duration of staying, it is visited once per day. The café is located there, thus it is possible to assume that the given employee spends lunch break there.

3.5. The Anomaly Heat Map View

The authors consider anomalies as irregular significant or insignificant deviations in the routes of the employees; and it is suggested to assess deviations within the group of employees or group of time intervals, e.g., day, that are characterized by similar routes. The deviations are shown using the heat map which is constructed in the following way. The Y-axis corresponds to the employees or days in the cluster; and the X-axis represents attributes of the vector generated from the log data as it is described above. Each element of the heat map represents distance of an attribute value of the sample from the cluster centroid. However, displaying the distances directly may produce a rather noisy picture on the one hand when the distances are comparable, or hide some deviations on the other hand, if distance variance is high. In addition, the significance of the deviation strongly depends on the attributes of the particular controlled zone. For example, a deviation of 20 min in the zone with offices cannot be compared with the similar deviation taken place in the zone with lifts.
The ranking mechanism consists of two stages: (1) calculation of the deviation ranking and (2) determination of threshold values for each controlled zone. The calculation of the deviation ranks is based on the calculation of the z-score reflecting how the current value of the attribute is above or low its mean value measured in number of standard deviations. The discretization approach used on the data preprocessing stage allows assessing deviations for a particular time interval.

The z-score takes values lying in the interval \([-4, 4]\) and indicates the proportion of the standard deviation the current value is greater or less than the average one. The values of the z-score in the range of \([-1.65, 1.65]\) represent the expected result, meaning that the behavior of the employees almost does not deviate from their typical behavior. Values exceeding the range \([-2.58, 2.58]\) indicate that at these points in time, employees demonstrate a non-standard behavior pattern that is not the result of a random process [41]. Thus, all deviations, having scores \([-4, -2.58]\) and \((2.58, 4]\) can be considered anomalous and need detailed investigation.

The threshold values set for each controlled zone determine what deviations would be displayed on the heat map. They implement filtering of the deviations that is based on time and spatial attributes of the deviations such as duration, the controlled zone the deviation took place and period of time during which the deviation occurred. Thus, an analyst may focus only on the deviations occurring in the interval from 8 to 12 am in the specific controlled zone, for example server room.

Figure 7 demonstrates the application of the proposed anomaly ranking mechanism to the assessment of the deviations in the employees’ routes.

Figure 7a shows the heat map that displays the z-scores calculated for the corresponding distances, the generated picture seems even more noisy, and this is explained by the fact that z-score in fact normalizes distances and highlights even minor ones, that is why new deviations with small z-scores appeared on the heat map.
The determination of threshold values for each controlled zone. The calculation of the deviation ranks is based on the calculation of the z-score reflecting how the current value of the attribute is above or below its mean value measured in number of standard deviations. The discretization approach used on the data preprocessing stage allows assessing deviations for a particular time interval. The z-score takes values lying in the interval \([-4, 4]\) and indicates the proportion of the standard deviation the current value is greater or less than the average one. The values of the z-score in the range of \([-1.65, 1.65]\) represent the expected result, meaning that the behavior of the employees almost does not deviate from their typical behavior. Values exceeding the range \([-2.58, 2.58]\) indicate that at these points in time, employees demonstrate a non-standard behavior pattern that is not the result of a random process [41]. Thus, all deviations, having scores \([-4, -2.58)\) and \((2.58, 4]\) can be considered anomalous and need detailed investigation.

The threshold values set for each controlled zone determine what deviations would be displayed on the heat map. They implement filtering of the deviations that is based on time and spatial attributes of the deviations such as duration, the controlled zone the deviation took place and period of time during which the deviation occurred. Thus, an analyst may focus only on the deviations occurring in the interval from 8 to 12 am in the specific controlled zone, for example server room.

Figure 7 demonstrates the application of the proposed anomaly ranking mechanism to the assessment of the deviations in the employees’ routes. The anomaly heat maps, displaying (a) distances between attribute values and corresponding values of the cluster centroid; (b) z-score of the distances; (c) z-score of the distances having values in the ranges \([-4, -2.58]\) and \((2.58, 4]\).

The result of filtering out all deviations which z-scores lies in the interval \((-2.58, 2.58)\) is shown in Figure 7c, and now it is clearly seen who exposes suspicious behavior, and the number of the days with suspicious routes decreased to the four.

4. An Approach to Analysis of the Employees’ Movement

To assess the applicability and evaluate the proposed approach, the authors developed the software prototype. Figure 8 depicts the main dashboard supporting suggested analysis workflow. It consists of 4 main views. The view B consists of 2 tabs. The SOM View tab contains two SOMs—the Employees SOM and the Periodicity SOM. Note, that instead of circles that mark SOM cells with clusters, the Periodicity SOM is equipped by the WeekCircle glyph that is also used to mark the presence of the cluster in the SOM cell. Its size like size of circle depends on the number of objects (days) in the cluster. The BandView tab displays raw data. The view C consists of two tabs—Pattern View and Anomaly View. The Pattern View shows the extracted route patterns, and Anomaly View
is a heat map of the possible anomalies in the employees’ routes. The Graph View (view D) shows the graph of visited controlled zones, and the properties of the analyzed objects—selected employee, group of co-workers or one particular visit of the selected zone—are presented by the Property View (view A).

![Figure 8. The analytical dashboard designed to support the proposed employees’ movement analysis.](image)

All visualization models used in views are interactive and interconnected. Clicking on each graphical element of all data visualization models, an analyst updates information displayed in the linked views. For example, the selection of the Employees SOM cell causes the update of data displayed in all other views: the Property View shows the members of the selected group, the Periodicity SOM displays groups of time periods that are characterized by similar employee routes, the Graph View and Pattern View present spatial and spatiotemporal patterns of co-workers’ trajectories, correspondingly, the BandView tab shows the routes of the selected employees, possible anomalies in the selected set of the routes are displayed in the Anomaly View. Selecting element from the Days SOM allows focusing on a particular set of days sharing similar movement pattern by displaying information in views A, C and D only for the selected dataset. Thus, the SOM-based views can be considered as graphical filters allowing analyst to reduce initial dataset and focus on areas of interest, and the heat map-based view serves as navigation panel used to spot possible anomalies. On the other hand, the BandView of the raw data and Anomaly View can be used to validate the output of the SOMs as too noisy heat map may indicate that the objects in the cluster are very diverse, and BandView shows the detailed information of the employees’ routes in the cluster. Authors suppose that these core interaction mechanisms between different visual models support analysis workflow proposed above.

Two datasets were used to assess the efficiency of the approach. Unfortunately, there is no publicly available dataset describing movement of staff of power station or other energy company as it is confidential information. That is why authors used one dataset artificially generated for the Mini-Challenge 2 of the VAST Challenge 2016, and another one that describes movement of the real employees of the middle-sized company located in the Saint-Petersburg, Russia, and that have some offices in the other cities of the Russia. Though these datasets do not describe the movement of the employees within critical infrastructure such as power station or other energy system, we consider that it is possible to test the efficiency of the approach using them. First of all, employees of power station or other energy companies may have more strictly defined routines and activities and as a
result their movement has more clearly defined patterns that are more easily to detect in comparison
to patterns present in the movement of staff belonging to companies of other types. For example,
usually software companies do not have strict working hours, the amount of working hours per week
is usually determined, thus employees may select a working schedule convenient for them. The next
two subsections describe the peculiarities of the datasets being analyzed, and obtained results of the
analysis using proposed technique.

4.1. The VAST Challenge 2016: Mini-Challenge 2 Use Case

The dataset provided within Mini-Challenge 2 of the VAST Challenge 2016 contains logs of the
proximity card readers that cover individual building zones. When an employee with proximity card
enters a new controlled zone, his/her card is detected and recorded. The dataset contains a two-week
set of logs. It should be noted that almost all areas are available to staff members even if they forget
their proximity cards. The proximity switch logs are complemented with building layout for the offices,
including the maps of the controlled zones, and the list of employees containing information on their
department and office assignments.

The proximity switch logs have the following format: <timestamp, type, prox-id, floor, zone>,
where timestamp stands for the time stamp of the log entry, type defines the type of the proximity card
reader, prox-id is the identifier of the employee, parameters floor and zone describes the location of the
controlled zone. To obtain a unique identifier of the controlled zone, the authors concatenated the
values of the floor and zone attributes.

The vectors describing the route of each employee were calculated as it is described in Section 3.1.
The duration $T$ of the period presented by the logs equals to 14 days, the authors set time interval
$\Delta t$ used to split it into sequence of time intervals equal to 4 h. In order to obtain vectors describing
employee’s movement, the number of visits and duration of staying of a particular employee were
calculated for each controlled zone and for each time interval and then concatenated into one vector.
Then these vectors were used as input for SOM clustering and proposed visualization techniques to
analyze the trajectories.

The analysis of the employees’ trajectories showed that they depend on the employees’ department
strongly. The representatives of some departments (Engineers, Information Technologies, Facilities,
Security, HR) have well-defined patterns in their routes that can be easily extracted. In this case,
the employees belonging to one department move alike, and existing difference can be explained by
the location of their work places. However, there are some departments (Administrative, Executive)
the representatives of which move rather chaotic, and the patterns are hardly could be determined
even for the days of week for one employee. For example, the employees belonging to the Engineers
department have rather similar routes that do not change during all periods of time. The number of
the department staff is 33, the following parameters for constructing the Employees’ SOM were set:
the size of the SOM was set to $10 \times 4$, number of training epochs equaled to 1000, the initial learning
rate was set to 0.9 and the learning function was inverse-of-time function to guarantee that all input
samples have approximately equal influence on the training result. Figure 9 shows the Employees
SOM constructed for this department. The circles mark SOM cells with clusters, and their size reflects
the number of employees in the cluster. The analysis of the SOM allows us to conclude that there are
7 clusters, two of them are rather numerous (clusters 1 and 2), while rest are small ones, consisting of
one person. It is also possible to conclude that the routes of employees belonging to clusters 1, 2, 4, 5,
6 and 7 significantly differ from the routes of employees belonging to cluster 3 as these clusters are
separated by a set of dark SOM cells. The color of cells surrounding clusters 1, 2, 4, 5, 6 and 7 is not
so dark, therefore it is possible to assume that the existing differences in the routes of corresponding
clusters are not so significant. All these assumptions were confirmed by detailed analysis of the data
presented by the BandView model.
The route of the employee of the cluster 7 is similar to the routes of the cluster 1, however he spends Energies department. The certain freedom in their movement is explained by their role in the organization as of the red stripes on the anomaly heat map and long band on the BandView.

The employee cwhelan regularly forgets to use his proximity card at the end of the work day that results in appearance of the sequence appearing shortly afterward for the slightly modified proximity id. The employee cwhelan lost his proximity card and he referred to the Security department to issue a new one. This fact is the best seen on the BandView visualization as an interrupted band for one proximity id and a new one.

The period of the employees belonging to the cluster 1 is similar to the routes of the cluster 1, however he spends his lunch time in the zone 2-1 where the café is located while his colleagues from the cluster 1 prefer going to the canteen located on the first floor (zone 1-2). The offices of the employees belonging to the clusters 4, 5 are located in two adjacent zones 2-6 and 2-7, correspondingly. Three engineers of the cluster 3 start their working day in the second half of the day approximately at 16:30, that is why they are separated from the others by the set of dark cells.

It turned out that employees belonging to the Engineers department spend most of their working time on their work places, and existing differences in the routes are explained by the location of their offices. The most numerous clusters (clusters 1, 2) have rather similar routes except that one group of employees spends their working time in the 2-1 zone, while another in the zone 2-2. The Periodicity SOM shows that their routes are very typical (Figure 10), their day patterns are presented in Figure 11.

The route of the employee of the cluster 7 is similar to the routes of the cluster 1, however he spends his lunch time in the zone 2-1 where the café is located while his colleagues from the cluster 1 prefer going to the canteen located on the first floor (zone 1-2). The offices of the employees belonging to the clusters 4, 5 are located in two adjacent zones 2-6 and 2-7, correspondingly. Three engineers of the cluster 3 start their working day in the second half of the day approximately at 16:30, that is why they are separated from the others by the set of dark cells.

The authors detected only few anomalies in the trajectories of the engineers. The employee tquiroz lost his proximity card and he referred to the Security department to issue a new one. This fact is the best seen on the BandView visualization as an interrupted band for one proximity id and a new one appearing shortly afterward for the slightly modified proximity id. The employee cwhelan regularly forgets to use his proximity card at the end of the work day that results in appearance of the sequence of the red stripes on the anomaly heat map and long band on the BandView.

The example of the department with employees having diverse trajectories is the Executive department. The certain freedom in their movement is explained by their role in the organization as
they perhaps do not have a strict time schedule. The single common feature of the routes is that they mostly spend time on the third floor where their offices are located. The diversity of their routes is confirmed by the results of the Employees SOM clustering: for the group of ten persons it detected six clusters, four of them contain one person and the rest consists of two employees.

The clustering of the days also indicated on the existing diversity as almost all days were classified as separated clusters. The main way to analyze such trajectories is to apply the BandView model to visualize the raw moves of the executive staff and the graph of the visited controlled zones to reflect statistics on the number of visits and duration of staying in each visited zone. These techniques were also helpful in detecting existing deviations in employees’ routes, as the produced heat map was too noisy to be interpreted due to variety of the trajectories or it was not possible to construct it for a cluster containing only one entity. Figure 12 shows the graph of the visited zones for the jsanjorje, the representative of the Executives, it is clearly seen that he visits zones of the third floor, spending most of the time in the zone 3-6 where his office located, he never visits offices of the second and first floors, and once was in the canteen of the zone 1-2 (Figure 13). The BandView model revealed one strange case when his proximity card was registered at midnight at the building entrance (zone 1-1).

**Figure 12.** The graph of visited zones by the employee of the Executive department.

**Figure 13.** The BandView visualization of the daily routes of the Executive department representative.

### 4.2. The Movement of the Employees of the Middle-Sized Company

The second dataset describes the activity of the employees of the real middle-sized company located in St. Petersburg. It was generated by the system for monitoring employees’ activity on the work places during one month. The proximity card readers are installed in all offices of the building, at the entrance and canteen. Due to specificity of the business processes adopted, the proximity card readers, monitoring some offices, were switched off during the period starting from 9.00 am till 18.00, meaning that it is not possible to form exact patterns of the movement during the work day, however, the proximity switches installed in the canteen, at entrance of the building and some offices functioned.
all day long. The logs contain events from different sources, including events from operating system, and have the following format: <employee id, data, location description, reserved, reserved, timestamp of the activity beginning, timestamp of the activity completion, duration, activity type, location address, data type, description, priority, flag>.

To apply the proposed approach, the data provided were transformed to the format necessary: <timestamp, employee-id, location-id>. The location-id was generated in the manner similar to the first use case. The time interval ∆t was set to 24 h as there is not much movement during the day.

The logs contain data about movement of the employees belonging to two departments—Managers and IT. The analysis showed that the trajectories of the employees depend on their department. The representatives of both departments have flexible schedule of the visits, but it was possible to conclude that there is a requirement to an obligatory number of working hours, as the employee may be absent on working place during work days and present at the days-off. The differences in the routes of Managers and IT employees are most clearly seen using the graph of the visited zones. Figure 14 shows the zones visited by the employees of these departments. The detailed results of the analysis of the IT employees’ routes are given below.

![Graph of the zones visited by the employees of (a) IT department; (b) Managers department.](image)

The number of the IT personnel is 6. The Employees SOM formed 5 clusters, all except one consisting of 1 person, indicating that their routes are diverse (Figure 15). The cluster 1 contains two employees, their typical routes are showed in Figure 16a. It is clearly seen that they work approximately 9 h each day and leave office 2 times per day for 20–30 min. The anomaly heat map revealed 6 days with severe deviations and 9 days with slightly atypical behavior (Figure 16b).

![The Employees SOM of the IT department.](image)
The insignificant deviations are explained by the changes in the number of visits of working place, while the duration of the staying inside the company is relatively the same. The severe deviations are associated with visiting atypical zones, forgetting using of the proximity card, and visiting the building in the atypical time. For example, the anomaly of the third day is explained by the presence of one employee in the work place during the day off (state holiday), the anomalies of the 11th and 15th days are the cases when the employee forgets to use proximity card when leaving the building. The visits of atypical zones—canteen (4-2) and offices (3-5, 3-9)—are registered on the 8th, 23th and 28th days. It should be noted, that the visits of the atypical zones are indicated by two bright stripes during the day, while other anomalies are characterized by one bright stripe, this is explained by that the changes in movement are registered for two zones. The days with anomalies also constitute separate clusters on the Periodicity SOM. The graphical representation of the anomalies using the BandView visualization is shown in Figure 17.

The cluster 2 consists of one employee only. His/her typical route looks very similar to the routes of the employees of the cluster 1 except that he/she was absent for 14 days during the period being analyzed.

The employee from the upper leftmost cluster 3 works 6 days per week (Figure 18). Every Monday he/she starts working early in the morning visiting the office 3-5 at the beginning of the work day, on the rest of the work days the employee works similar to the employees from the clusters 1, 2. On Saturdays, he/she works in the office 3-5 in the second part of the day. The following anomalies were registered for this employee: once he/she worked on Sunday for one hour, and twice he visited zone 3–9 atypical for her/him.
5. The Approach Efficiency Evaluation

In the general case, the proposed approach to the analysis of the employees’ trajectories within building allows detecting the following types of anomalies:

1. The employee loses his/her proximity card (he/she may restore it on the same day, on the next day; two proximity cards are used simultaneously);
2. The employee forgets to use his/her proximity card;
3. The employee spends atypical period of time in the typical zone (atypical duration may be too long, too short; it may be employee coming to work at the weekend, atypical duration may be caused by missing visit to a controlled zone at a given time interval);
4. The employee visits atypical zone (atypical duration may be too long, too short; the employee leaves the building or is absent the whole day);
5. Two employees sharing one proximity card.

Each type of the anomaly is characterized by different visual patterns for each visualization model used. Some of anomalies form specific pattern on the anomaly heat map, some are noticeable already on the SOM views as separate clusters, while others are clearly seen only on the BandView visualization.

Table 1 shows graphical patterns indicating about presence of the anomaly.

| Graphical Signs of the Anomaly |
|--------------------------------|
| The Employees SOM |
| A circle, corresponding to a suspicious employee, is nearby the circle of greater size |
| The employee loses his/her proximity card |
| - Two proximity cards are used simultaneously |
| The employee spends atypical period of time in the typical zone |
| - The employee comes to work at the weekend |
Table 1. Cont.

| Graphical Signs of the Anomaly | Type of the Anomaly and Description |
|-------------------------------|-------------------------------------|
| The Days SOM                  |                                     |
| A circle, corresponding to suspicious day (days) employee, is nearby the circle of greater size. The WeekCircle glyph of a circle shows days similar to days of WeekCircle glyph of a larger circle. The longer duration of misbehavior the more chances that the day is classified as a separate cluster. | The employee loses his/her proximity card |
|                               | - The employee restores it on the same day, on the next day |
|                               | - Two proximity cards are used simultaneously |
| Two employees using one proximity card simultaneously | |
| A circle, corresponding to suspicious day (days) employee, is nearby the circle of greater size. The WeekCircle glyph of a circle shows days-off while WeekCircle glyph of a larger circle shows working days. Or The WeekCircle glyph of circle on a SOM map corresponding to a suspicious group of days shows a day-off among working days. | The employee spends atypical period of time in the typical zone |
|                               | - Employee does not visit controlled zone at a given time interval, causing time shifts in durations of staying in the other zones |
|                               | - The employee forgets to use his/her proximity card (longer than usual duration of staying in the typical zone) |
|                               | - The employee leaves the building |
| The employee visits atypical zone | - The duration of staying is long |
| The day will constitutes a separated cluster or is assigned to the cluster of day-offs in case of their presence | The employee spends atypical period of time in the typical zone |
|                               | - The employee comes to work at the weekend |
| The Anomaly heat map          |                                     |
| The sequence of bright stripes in the zones typical for the employee. The longer duration of anomaly the more saturated color of corresponding stripes. | The employee loses his/her proximity card |
|                               | - The employee restores it on the same day, on the next day |
| Two proximity cards are used simultaneously | |
| The employee spends atypical period of time in the typical zone | |
|                               | - The employee forgets to use his/her proximity card (longer than usual duration of staying in the typical zone) |
|                               | - Employee does not visit controlled zone at a given time interval, causing time shifts in durations of staying in the other zones |
|                               | - The employee spends less time than usual in the typical zone (shorter duration of staying in the typical zone) |
|                               | - The employee is absent the whole day |

| Graphical Signs of the Anomaly | Type of the Anomaly and Description |
|--------------------------------|-----------------------------------|
| Bright stripe in the zone with entrance during given period of time, and a sequence of lighter stripes in the adjacent controlled zones appear | The employee spends atypical period of time in the typical zone |
| Atypical location of dark stripes | The employee leaves the building |
| Bright stripe in the atypical location, and a set of stripes in the typical zones appear if a duration of anomaly is long | The employee comes to work at the weekend |
| The sequence of bright stripes in the zones typical for each employee. | Two employees using one proximity card simultaneously |
| The BandView model | |
| Two proximity card identifiers belonging to one employee are present, a gap between two bands corresponding to the route of the employee exists | The employee loses his/her proximity card |
| Two proximity cards are used simultaneously | The employee restores it on the same day, on the next day |
| Two proximity card identifiers belonging to one employee are present, and two different or similar routes take place at the same period of time | Two proximity cards are used simultaneously |
| A number of the band segments corresponding to one employee is present | Two employees using one proximity card simultaneously |
| A colored band of the employee and empty rows corresponding to his/her co-workers | The employee spends atypical period of time in the typical zone |
| A colored band of the employee and empty rows corresponding to his/her co-workers | The employee comes to work at the weekend |
| A gap in the sequence of the colored band segments is present | The employee leaves the building |
| Empty row corresponding to the route of the employee is present | The employee is absent the whole day |
| A new segment of the band appears. The lengths of segments corresponding to employees’ typical zones decreases | The employee visits atypical zone |
Interestingly, the most informative visualization model is BandView, it clearly shows almost all types of anomalies. It is also clearly seen that duration of the anomaly defines a type of visual pattern, if it is long and its duration exceeds the duration of the time interval $\Delta t$ selected on the data preprocessing stage, the day with anomalous route or even the employee constitutes a separated cluster on the corresponding SOM view. If the anomaly duration is less than $\Delta t$ but takes place at the atypical zone it is clearly seen on the anomaly heat map as a bright stripe. The anomalies in the typical zones with duration less than $\Delta t$ are the most difficult to detected, they are clearly seen only when their duration is more then $1/12 \Delta t$. The simultaneous usage of two card readers and usage of one card by two employees causes increase of activity in the typical zones and therefore are clearly seen on the heat map. It is worth noticing that some anomalies have similar visual patterns and only the BandView visualization may provide detailed description on the type of the detected anomaly.

As the proposed approach utilizes tightly connected data mining and visualization techniques, it is possible to evaluate the efficiency of the selected analysis models, i.e., the appropriateness of the data preprocessing step, results of the SOM clustering and detection of the anomalies using statistical assessment and the efficiency of the visualization models, i.e., the usability of the graphical presentation of the SOM, the BandView model and the heat map.

To assess the accuracy of the analysis models, the authors chose the dataset provided within the VAST Challenge 2016 as the solutions to the challenge could be used as a ground proof to the detected anomalies. According to the answer sheet provided, there are 4 types of the anomalies: (1) visiting typical zone in the atypical daytime (weekends or midnight), (2) forgetting using card at the end of the work day, (3) multiple card issues and (4) simultaneous usage of cards linked to one employee. The authors also manually analyzed the dataset, and discovered some minor deviations in the employee’s routes connected with the atypical duration of staying in the typical controlled zones, forgetting using a card during the work day. The following parameters of the automated analysis models were used during assessment process: $\Delta t = 4$ h at the preprocessing step; the size of the SOM net is defined as $5 \times \sqrt{n}$, where $n$ is the number of vectors (employees or days), the SOM learning rate function is inversely proportional to time, and initial learning rate is 0.9. The anomalies were detected using heat maps displaying or z-score of the distances of the objects from cluster centroids. Figure 19 shows overall number of anomalies, their distribution by the types and the rate of the detected anomalies using the proposed approach. It should be noted that all false positive results were associated with the first type of anomaly—visiting of typical zone with anomalous duration of staying.

To assess the effectiveness of the visualization techniques, the authors applied the inspection method which is widely used to evaluate multidimensional data visualization techniques. This approach assumes usage of benchmark datasets and group of experts that examine the effectiveness of the proposed visualization technique to solve different tasks by analyzing visually the graphical output. To measure effectiveness, the task completion metric was used, it is calculated as the proportion of the correctly completed tasks.

![Figure 19](image_url)  
*Figure 19. The rate of detected anomalies of different type.*
The authors used the VAST Challenge 2016 dataset to form a benchmark dataset, it was split into several test datasets containing routes of 10 employees belonging to one or several departments. We also developed the questionnaire that consisted of two parts: the first part of questions is aimed to assess the functional characteristics of the proposed tool, and the second is aimed to the level of subjective effectiveness of the proposed visualization techniques.

The first part of the questionnaire contained the following questions:

1. Determine the number of groups of employees, having similar routes; specify their capacity and composition;
2. Describe the typical route of the employees belonging to the groups;
3. Describe the main differences in the routes of employees belonging to different groups;
4. Determine if the employees’ routes depend on the day of the week (for each group of employees); if detected, describe the existing periodicity in format (day of week, time, duration and controlled zone or sequence of the controlled zones);
5. Determine if there are anomalies in the routes of employees (for each group of employees), if detected, describe them in format: employee ID, day, time, duration, controlled zone (or sequence of the controlled zones). Try to explain the nature of the anomaly (1-2 sentences).

The second part of the test consisted of questions aimed to evaluate the simplicity and clarity of visualization models, methods of interaction with them. When answering them, it was necessary to rate the models from 1 to 5, where 1 corresponds to very easy, and 5—very difficult.

Some examples of the questions are listed below:

1. Was it easy or difficult to set up the parameters for the Employees/Periodicity SOM (1—it is easy to set up the parameters, the clustering results are correct; 2—it is more difficult to set up the parameters, but the clustering results are correct; 3—it is not easy to set up the parameters, but the clustering results are valid, there is a cluster that contains employees with minor differences in the routes; 4—setting up the parameters is difficult, the results of clustering are acceptable, in several clusters there are 1-2 employees with significant differences in the movement; 5—it is difficult to choose the parameters, the results of clustering are not acceptable)?
2. Was it easy or difficult to interpret the graphical output of the Employees/PeriodicitySOM?
3. Was it easy or difficult to interpret the output of the BandView visualization model?
4. Was it easy or difficult to interpret the graph of controlled zones?
5. Was it easy to interpret the output of the heatmap visualization (1—yes, 2—yes, all anomalies are clearly seen, however interpretation takes time; 3—the anomalies are clearly seen, but insignificant ones are “lost” among noise, the thorough analysis of the routes is required; 4—there is much noise, however the significant anomalies are clearly seen, though their analysis requires much effort and time; 5—there is much noise, the implemented filtering mechanisms do not allow detection of the anomalies) ?
6. Rate the choice of the color scheme for the visualization models.

The focus group consisted of 5 specialists having experience in information security and intrusion detection techniques, 2 specialists in data analysis, and 10 graduate students studying information security. Among specialists there were both practitioners and scientific researchers. Before conducting evaluation process the invited experts were given a brief introductory tutorial about the proposed approach, visualization and analysis models, and the tool; the authors discussed the peculiarities of the SOM clustering and recommended parameters, how other visualization models are used to support analysis process. Then the participants were given a description of the Mini challenge 2 of the VAST Challenge 2016 and had a possibility to ask questions if they occurred. Then they were given one of prepared datasets each and were asked to fill in a questionnaire. After the completion of the analysis tasks the participants were also asked to give a feedback about the tool by making suggestions or critics.
The task completion for the first part of the questionnaire of tasks was 93% for specialists and 84% for students and postgraduates. The mean rate for the simplicity and clarity of visualization models was 2.1.

Almost all participants correctly determined the number of the employees having similar routes and provided correct description of their typical routes. There were some minor mistakes when employees with light differences in the routes (they spent several hours in the first part of the day in different but neighboring zones) were assigned to one cluster, and the deviations in the route of the one employee despite of their regular character were classified as anomalous. Lately, when discussed the problem with the participants, the authors found out that the cause of the error was the incorrect selection of the Employees’ SOM size which were set equal to the number of employees. All significant anomalies in the employees’ routes were such as visiting of the atypical zones, working at the weekends, forgetting using proximity card at the end of the day and cases of multiple issues of the proximity cards. The majority of the mistakes were associated with detection of the visits of typical zones with atypical duration during the work day. The main causes of these errors were the wrong selection of the time interval at the data preprocessing step and setting up wrong parameters of the SOM. For example, some participants missed deviations up to 40 min when they set up time interval equal to 8 h. The wrong selection of the SOM size also “hid” anomalies producing very noisy heat map of anomalies; and the process of the anomaly detection came down to the analysis of the BandView model.

The visualization models were ranked rather high, and many experts highlighted that they allow forming comprehensive understanding of the data highlighting both route patterns and anomalies. The participants marked that the Employees SOM clearly shows possible similarities between clusters, and the heat map can be used to validate the results of the clustering: too noisy heat map may indicate about wrong choice of the initial parameters of the SOM. They also liked the BandView model, it depicts the raw data clearly and is easy to be interpreted, and moreover it is helpful in understanding possible interaction between co-workers. To make the usage of the tool more effective, the experts advised developing interaction mechanisms that magnify the heat map areas and coordinate magnified areas with the BandView to enhance the process of anomaly detection. They also recommended adding a possibility to form a formal description of the patterns and anomalies detected in order to implement for search-by-example interaction mechanism or use it in training automated analysis models.

The analysis of the related works showed that there is not much work in the analysis of the proximity logs, however, the similar problem is investigated in [32] and [42]. The solutions of the VAST Challenge 2016 could be also considered as the approaches aimed to analyze employees’ routes, that is why we included solutions of City University London and the KU Leuven University awarded for the Outstanding Presentation of Patterns in Context and Strong Support for the Anomaly Detection correspondingly in the comparison process. Considering the questions formulated by the experts in Section 3, the ability to detect patterns and anomalies in the movement are the key requirements to the analysis technique.

To compare the proposed approach with tools presented in the open access literature, the following comparison criteria were determined:

C1: requirements to the source data;
C2: ability to detect groups of employees having similar behavior;
C3: ability to present patterns in the employees’ movement;
C4: ability to detect anomalies in the employee’s movement;
C5: ability to detect periodicity in the employees’ movement;
C6: ability to present interaction between co-workers.

The solutions of the VAST Challenge 2016 by City, University London and the KU Leuven University and one described in [32] are visualization-driven approaches. The approach suggested by City, University London is rather close to the suggested one. The graphical representation of the raw data is implemented by the visualization model similar to the BandView model, the color is used to
encode either attributes of the controlled zones or indicate whether the employee office is located in the zone. To detect groups of the employees with similar routes the author used KMeans++ algorithm [43]. The clustering algorithm is applied to the sequence of zones classified by either their attributes or location of the employee’s office. It is not clear, whether the temporal attributes of the moves are taken into the consideration. Though the visualization displaying raw data shows all employees, they are grouped according to the clustering results, thus the groups of employees having similar behavior are seen. The tool implements various interaction mechanisms that support analysis process including possibility to make notes during analysis, however, the detection of anomalies is done by manipulating raw data, there are no automatic means to highlight deviation from the typical routes.

The solution provided by KU Leuven University uses the Gantt chart to visualize raw data, not very compact to present data for several employees. The patterns of the visits for each controlled zone are displayed using step charts, but they do not provide information about typical routes of the employees. However, they suggested an interesting approach to detect anomalies in the proximity dataset. The authors calculate sequence score that measures the monotony level in the movements of the particular employee and time score that assesses the duration of staying in the particular location based on all available data. The obtained scores are visualized using scatter plots constructed for each day for each employee.

The approach described in [32] is also targeted to explore the routes of employees and was tested using the VAST Challenge dataset. The suggested Event Quiltmap visualization model resembles much to the BandView visualization model and one presented by City, University London except that the vertical axis corresponds to the time. The groups of the employees are detected using clustering algorithm that assesses distances between each pair of employees over all periods of time being analyzed. Like the City, University London solution, the employees are grouped using obtained clusters. No specific means for anomaly detection are suggested.

Thus, it is possible to conclude the following. None of the approaches described above allows one to reduce dimension of the input data by forming typical routes for the groups of employees with similar behavior as well as to detect the possible dependence of the routes on the days of week. The application of the filtering mechanism is the only possibility to focus on the particular employee or group of employees. The periodicity in the movement can be established only by examining the Gantt chart-based visualization models. The special support for automated detection of the anomalies is developed in the KU Leuven University solution. The temporal deviations in the employee’s daily routes are determined on the basis of the all available data, but not in the context of the detected group as in the approach described in the paper, thus it is possible to conclude that certain types of anomalies could be missed as different groups of employees move and stay in the controlled zone differently.

The approach presented in [42] does not use visualization techniques to analyze employees’ trajectories. The authors define the notion of the movement motif that is based on the controlled zone’ characteristics, employees’ position in the organization, and temporal attributes. This allowed authors to detect high level activities such as staying in the office, going for lunch, visiting meeting, etc. The typical behavior of the employee is described in form of detected activities and could be represented in form of Petri Net. The moves are considered anomalous if it is not possible to determine any motif for them. The efficiency of the approach described in [42] is comparable to the efficiency of the one presented in the paper.

None of the approaches described above were tested against real world data, while the approach proposed by the authors showed good results in analysis of the data obtained from the real world company, it was able to determine both groups of employees with similar routes, existing periodicity in the movement and highlight deviations in their movement. However, it should be noted that all approaches including the one proposed in the paper do not allow forming a formal description of the typical routes and anomalies that could be used in detection of the anomalies in real time mode.

Table 2 summarizes the results of the comparison of the suggested approach to the described above. We put “±” for those comparison criteria (tasks) that could be solved using the Gantt chart-based
visualization model, i.e., BandView model, Event Quiltmaps as it is a very helpful way for representing raw data about employees moves and could be used in establishing interaction patterns between employees and their typical routes.

Table 2. Results of the comparison of the suggested approach with other tools.

| Criteria | [31] | [41] | KU Leuven University | City, University London | The Proposed Approach |
|----------|------|------|----------------------|-------------------------|----------------------|
| C1       | (1) Logs from access control systems | (1) Logs from access control systems | (1) Logs from access control systems | (1) Logs from access control systems | (1) Logs from access control systems |
|          | (2) Floor plan | (2) Floor plan with description of the zone features | (2) Floor plan with description of the zone features | (2) Floor plan with description of the zone features | (2) Floor plan with description of the zone features |
|          | (3) List of employees with their positions | (3) List of employees with their positions | (3) List of employees with their positions | (3) List of employees with their positions | (3) List of employees with their positions |
| C2       | +    | −    | +                   | +                       | +                     |
| C3       | +    | +    | +                   | +                       | +                     |
| C4       | −    | +    | −                   | −                       | −                     |
| C5       | ±    | −    | ±                   | ±                       | ±                     |

Thus, it is possible to conclude that the approach suggested in the paper is effectively used in the exploratory analysis of the employees’ trajectories. The experiments showed that it supports answering the questions and requirements stated by security analysts, as presented in Section 3, in following way.

1. To detect groups of employees sharing similar behavior, the SOM Employees is used; the detailed information on employees including their position is presented by Property View and BandView model visualizing raw data.

2. The typical daily route of employees belonging to one group is presented using BandView model in Pattern View, the existing dependency of route from week day is defined by the Periodicity SOM and is displayed by special glyph WeekCircle.

3. Any existing interactions between co-workers could be analyzed using BandView model, implemented filtering interaction mechanism allows focusing on particular department, time interval, controlled zone.

4. Different types of anomalous behavior of employees could be detected using different visualization models—Employees SOM and Periodicity SOM, anomaly heat map and BandView model. Their signs are given in the Table 1.

5. The information about detected anomaly is provided by BandView model visualizing raw data, it allows answering the question when and where the anomaly took place, who else was there simultaneously.

If included in the user behavior analysis module of the security information and event management system, it may provide better understanding on how people behave in the organization.

However, the efficiency evaluation process highlighted the most critical issues of the approach suggested—setting up the initial SOM parameters. This problem is more complicated as it requires end user to understand how input parameters influence on the SOM clustering.

Currently, the authors suggest using default parameters as they are defined in Usage Scenario section and were used during evaluation process, however the setting up process could be simplified to the end user if it is supported by an extra visualization models projecting multidimensional space of the source data into the two-dimensional one. This problem defines the direction of the enhancement of the approach presented.

6. Conclusions

Usage of advanced technological solutions in the energy sector on the one hand increased the efficiency of the energy consumption, environmental protection, but on the other hand made it vulnerable to different types of cyber threats that may result in severe impact on all aspects of human living. Insider threat detection is one of the most complicated tasks of cyber security. Monitoring
employees’ movement allows detection of potentially dangerous situations that could be caused by either careless attitude to job responsibilities or intended malicious activities. The analysis of the employees’ routes may provide better insight on existing business and technological routines, access restrictions and assist in monitoring the compliance to the established ones. In fact, this task is relevant to provision of energy organization physical and cyber security and impacts its sustainability in the whole.

The usage of the approach proposed may significantly increase the efficiency of the access control logs analysis as it reveals the groups of employees having similar routes and shows deviations in their routes. The groups of employees with similar behavior are detected using the interactive self-organizing Kohonen maps that take into account both temporal and spatial attributes of their movement, the route patterns are displayed using the Gannt chart based visualization model and the graph of the visited controlled zones. The deviations in employees’ trajectories are also assessed both in spatial and temporal context and are presented using heat map. The authors implemented the special anomaly ranking mechanism in order to simplify the process of the anomaly detection.

Thus, the approach suggested is aimed to detect anomalies on employee routes that could be caused by both real threat or carelessness of an employee, as anomaly is considered as deviation from some typical behavior. The approach allows describing the anomaly detected—when and where it took place, how long it lasted and who else was in the same zone, and the analyst could make hypothesis of what is happening—a serious violation of the security and safety policy or some routine or benign mistake of the employee, like pressing the wrong button in the lift cabin (without leaving it on wrong floor), taking into account these facts. However, it should be noted that the insider threat detection is one of the most complicated tasks in cyber security that requires considering many issues including psychological state of the employee; and the described approach provides insight only on aspects relating to employees’ movement.

The proposed approach was tested using artificially generated and real world data. The results of the efficiency evaluation process are presented in the paper. They showed that the accuracy of the approach in detection of the anomalies in employees’ routes is rather high, and the visualization models clearly convey all information. They also defined the direction of future research work connected with the elaboration of recommendations to the setting up the analysis model parameters.

Author Contributions: Methodology, E.N. and I.K.; software, E.N., I.M.; validation, E.N., I.M.; investigation, E.N., I.M.; resources, I.K.; writing—original draft preparation, E.N., I.M.; writing—review and editing, E.N., I.K.; supervision, I.K. All authors have read and agreed to the published version of the manuscript.

Funding: Research is carried out with support of Ministry of Education and Science of the Russian Federation as part of Agreement No. 05.607.21.0322 (identifier RFMEFI60719X0322).

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Cassotta, S.; Sidortsov, R. Sustainable cybersecurity? Rethinking approaches to protecting energy infrastructure in the European High North. Energy Res. Soc. Sci. 2019, 51, 129–133. [CrossRef]
2. Enhancing Canada’s Critical Infrastructure Resilience to Insider Risk. Available online: https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/nhnctg-rtrct-nfrstrctr/index-en.aspx (accessed on 12 June 2020).
3. Tan, L.; Hu, M.; Lin, H. Agent-based simulation of building evacuation: Combining human behavior with predictable spatial accessibility in a fire emergency. Int. J. Inf. Sci. 2015, 295, 53–66. [CrossRef]
4. Rawassizadeh, R.; Momeni, E.; Dobbins, C.; Gharibshah, J.; Pazzani, M. Scalable daily human behavioral pattern mining from multivariate temporal data. IEEE Trans. Knowl. Data Eng. 2016, 28, 3098–3112. [CrossRef]
5. Wang, J.; Gupta, M.; Rao, H.R. Insider Threats in a Financial Institution: Analysis of Attack-Proneness of Information Systems Applications. MIS Q. 2015, 39, 91–112. [CrossRef]
6. Kotenko, I.; Stepashkin, M.; Doynikova, E. Security analysis of information systems taking into account social engineering attacks. In Proceedings of the 19th International Euromicro Conference on Parallel, Distributed, and Network-Based Processing, PDP 2011, Ayia Napa, Cyprus, 9–11 February 2011; Volume 5739056, pp. 611–618. [CrossRef]

7. Hubstaff Employee Monitoring Software. Available online: https://hubstaff.com/employee_monitoring (accessed on 29 August 2019).

8. Observe, I.T. Insider Threat Solution. Available online: https://www.observeit.com/insider-threat-solution (accessed on 29 August 2019).

9. Suprema Access Control Solutions. Available online: https://www.supremainc.com/en/Solutions/AccessControl-TimeandAttendance (accessed on 7 September 2019).

10. Bussa, T.; Litan, A.; Phillips, T. Market Guide for User and Entity Behavior Analytics. Gartner Research ID G00292503. 2016. Available online: https://www.gartner.com/doc/3538217/market-guide-user-entity-behavior (accessed on 7 September 2019).

11. Novikova, E.; Murenin, I. Visualization-Driven Approach to Anomaly Detection in the Movement of Critical Infrastructure. Lecture Notes in Computer Science, 10446; Springer: Cham, Switzerland, 2017; pp. 50–61. [CrossRef]

12. Novikova, E.S.; Murenin, I.N.; Shorov, A.V. Visualizing anomalous activity in the movement of critical infrastructure employees. In Proceedings of the 2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), St. Petersburg, Russia, 1–3 February 2017; pp. 504–509. [CrossRef]

13. Novikova, E.S.; Murenin, I.N. The Technique of the Visual Analysis of the Organization Employees Routes For Anomaly Detection. SPIIRAS Proc. 2017, 5, 57–83. [CrossRef]

14. Kisleiveich, S.; Mansmann, F.; Nanni, M.; Rinzivillo, S. Spatio-Temporal Clustering: A Survey Data Mining and Knowledge Discovery Handbook; Maimon, O., Rokach, L., Eds.; Springer: Boston, MA, USA, 2010; pp. 855–874. [CrossRef]

15. Cabanes, G.; Bennani, Y.; Fresneau, D. Mining RFID Behavior Data using Unsupervised Learning. Int. J. Appl. Logist. 2010, 1, 28–47. [CrossRef]

16. Demšar, U.; Buchin, K.; Cagnacci, F.; Safi, K.; Speckmann, B.; van de Weghe, N.; Weibel, R. Analysis and visualisation of movement: An interdisciplinary review. Mov. Ecol. 2015, 3, 1–24. [CrossRef]

17. Morris, B.; Trivedi, M. Learning trajectory patterns by clustering: Experimental studies and comparative evaluation. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 312–331. [CrossRef]

18. Schreck, T.; Bernard, J.; von Langensberger, T.; Kohlhammer, J. Visual cluster analysis of trajectory data with interactive Kohonen maps. Inf. Vis. 2009, 8, 14–29. [CrossRef]

19. Andrienko, G.; Andrienko, N.; Bak, P.; Bremm, S.; Keim, D.; Landesberger, T.; Politz, C.; Schreck, T.A. A framework for using self-organising maps to analyse spatio-temporal patterns, exemplified by analysis of mobile phone usage. J. Locat. Based Serv. 2010, 4, 3–4. [CrossRef]

20. Novikova, E.; Kotenko, I. Analytical Visualization Techniques for Security Information and Event Management. In Proceedings of the 21st Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, Belfast, UK, 27 February–1 March 2013; pp. 519–525. [CrossRef]

21. Andrienko, N.; Andrienko, G. Visual analytics of movement: An overview of methods, tools and procedures. Inf. Vis. 2013, 12, 3–24. [CrossRef]

22. Andrienko, G.; Andrienko, N.; Hurter, C.; Rinzivillo, S.; Wrobel, S. Scalable Analysis of Movement Data for Extracting and Exploring Significant Places. IEEE Trans. Vis. Comput. Graph. 2013, 19, 1078–1094. [CrossRef] [PubMed]

23. Abel, J.; Sander, N. Quantifying Global International Migration Flows. Science 2014, 343, 1520–1522. [CrossRef] [PubMed]

24. Lu, M.; Lai, C.; Ye, T.; Liang, J.; Yuan, X. Visual Analysis of Route Choice Behaviour based on GPS Trajectories. In Proceedings of the 2015 IEEE Conference on Visual Analytics Science and Technology (VAST), Chicago, IL, USA, 25–30 October 2015; pp. 203–204. [CrossRef]

25. Liu, H.; Gao, Y.; Lu, L.; Liu, S.; Qu, H.; Ni, L.M. Visual analysis of route diversity. In Proceedings of the 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), Providence, RI, USA, 23–28 October 2011; pp. 171–180. [CrossRef]
26. Michael, S.; Robert, K.; Rolando, G.; Xing, L.; Ross, M. A Visual Analytics Framework for Exploring Theme Park Dynamics. *ACM Trans. Interact. Intell. Syst.* 2018, 8, 1–27. [CrossRef]

27. Krüger, R.; Thom, D.; Ertl, T. Semantic Enrichment of Movement Behavior with Foursquare-A Visual Analytics Approach. *IEEE Trans. Vis. Comput. Graph.* 2015, 21, 903–915. [CrossRef][PubMed]

28. Von Landesberger, T.; Brodkorb, F.; Roskosch, P.; Andrienko, N.; Andrienko, G.; Kerren, A. Mobilitygraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering. *IEEE Trans. Vis. Comput. Graph.* 2016, 22, 11–20. [CrossRef][PubMed]

29. Andrienko, G.; Andrienko, N.; Schumann, H.; Tominski, C. Visualization of Trajectory Attributes in Space–Time Cube and Trajectory Wall. *Lecture Notes in Geoinformation and Cartography;* Springer: Berlin/Heidelberg, Germany, 2014; pp. 157–163. [CrossRef]

30. Guo, C.; Xu, S.; Yu, J.; Zhang, H.; Wang, Q.; Xia, J.; Zhang, J.; Chen, Y.; Qian, Z.; Wang, C.; et al. Dodeca-Rings Map: Interactively Finding Patterns and Events in Large Geo-temporal Data. In Proceedings of the 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), Paris, France, 9–14 October 2014; pp. 353–354. [CrossRef]

31. Gupta, S.; Dumas, M.; McGuffin, M.; Kapler, T. MovementSlicer: Better Gantt charts for visualizing behaviors and meetings in movement data. In Proceedings of the 2016 IEEE Pacific Visualization Symposium (PacificVis), Taipei, 19–22 April 2016; pp. 168–175. [CrossRef]

32. McNamara, D.; Tapia, J.; Ma, C.; Luciani, T. Spatial Analysis of Employee Safety Using Organizable Event Quiltmaps. In Proceedings of the IEEE VIS 2016 Workshop on Temporal and Sequential Event Analysis, Baltimore, MD, USA, 24 October 2016.

33. Yen, P.Y.; Kelley, M.; Lopetegui, M.; Saha, A.; Loversidge, J.; Chipps, E.; Gallagher-Ford, L.; Buck, J. Nurses’ Time Allocation and Multitasking of Nursing Activities: A Time Motion Study. *AMIA Annu. Symp.* 2018, 2018, 1137–1146.

34. Shneiderman, B. Dynamic queries for visual information seeking. *IEEE Softw.* 1994, 11, 70–77. [CrossRef]

35. WaveTrend Access Control. Available online: http://www.wavetrend.net/access-control.php (accessed on 2 September 2019).

36. Lovrić, M.; Milanović, M.; Stamenković, M. Algorithmic methods for segmentation of time series: An overview. *J. Contemp. Econ. Bus. Issues* 2014, 1, 31–53.

37. Liu, S.; Maljovec, D.; Wang, B.; Bremer, P.; Pascucci, V. Visualizing High-Dimensional Data: Advances in the Past Decade. *IEEE Trans. Vis. Comput. Graph.* 2017, 23, 1249–1268. [CrossRef]

38. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* 1982, 43, 59–69. [CrossRef]

39. Ultsch, A. Self-organizing neural networks for visualization and classification. In *Information and Classification;* Springer: Berlin/Heidelberg, Germany, 1993; pp. 307–313. [CrossRef]

40. Brewer, C.A. Color use guidelines for mapping and visualization. *Mod. Cartogr.* 1994, 2, 123–147. [CrossRef]

41. Caldas de Castro, M.; Singer, B. Controlling the False Discovery Rate: A New Application to Account for Multiple and Dependent Test in Local Statistics of Spatial Association. *Geogr. Anal.* 2006, 38, 180–208. [CrossRef]

42. Novikova, E.; Bekeneva, Y.; Shorov, A. The Motif-Based Approach to the Analysis of the Employee Trajectories within Organization. *Secur. Commun. Netw.* 2018, 2018, 1–12. [CrossRef]

43. Arthur, D.; Vassilvitskii, S. K-means+: The advantages of careful seeding. In Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA ’07). Society for Industrial and Applied Mathematics, New Orleans, LA, USA, 7–9 January 2007; pp. 1027–1035.