Chapter 13 - Visual Analytics Methods for Movement Data

Gennady Andrienko[1], Natalia Andrienko[1], Ioannis Kopanakis[2], Arend Ligtenberg[3], and Stefan Wrobel[1]

[1] Fraunhofer Institute IAIS, Sankt Augustin, Germany {gennady.andrienko, natalia.andrienko, stefan.wrobel} @iais.fraunhofer.de
[2] Technological Educational Institute of Crete, Greece i.kopanakis@emark.teicrete.gr
[3] Wageningen UR, Centre for GeoInformation Arend.Ligtenberg@wur.nl

Summary. The chapter considers the use of interactive visual techniques for detection of various patterns and relationships in movement data. While a variety of techniques and tools for visualisation of movement data exist, very few of them are applicable to massive data. Moreover, movement data have quite a complex structure, which necessitates the use of multiple heterogeneous displays showing different aspects of the data. Linking between the displays is essential; however, traditional techniques for display linking are not scalable to large data collections.

The chapter outlines a roadmap to the creation of visualisation-centred tools for analysis of massive movement data. Besides visual representation of data, the tools involve data aggregation and other transformations. All the suggested visualisation techniques display data in an aggregated form and are therefore scalable. There are open issues that require further research. One of them is how several displays with differently aggregated data can be effectively linked while individual data items are not available in the computer memory. Another problem is visualisation of discovered patterns for the purposes of knowledge synthesis and communication.

1 Introduction

All the power of computational techniques for data processing and analysis is worthless without human analysts choosing appropriate methods depending on data characteristics, setting parameters and controlling the work of the methods, interpreting results obtained, understanding what to do next, reasoning and drawing conclusions. To enable effective work of human analysts, relevant information must be presented to them in an adequate way. Since visual representation of information greatly promotes human's perception and cognition, visual displays of data and results of computational processing play a very important role in analysis.

However, a simple combination of visualisation with computational analysis is not sufficient. The challenge is to build analytical tools and environments where the power of computational methods is synergistically combined with human's background knowledge, flexible thinking, imagination, and capacity for insight. This is the main goal of the emerging multidisciplinary research field of Visual Analytics (Thomas and Cook [45]), which is defined as the science of analytical reasoning facilitated by interactive visual interfaces.

Analysis of movement data is an appropriate target for a synergy of diverse technologies including visualisation, computations, database queries, data transformations, and other computer-based operations. In this chapter, we try to define what combination of visual and computational techniques can support the analysis of massive movement data and how these techniques should interact. Before that, we shall briefly overview the existing computer-based tools and techniques for visual analysis of movement data.

2 State of the art

2.1 Visualisation fundamentals

In a strict sense, visualisation is representation of data in a visual form, i.e. creating various pictures from data: graphs, plots, diagrams, maps, etc. For this purpose, items of data are translated into graphical features, such as positions within a display, colours, sizes, or shapes. For the visualisation to be effective, the translation is done according to the established principles and rules (see, for example, Bertin [6] or a summary in Andrienko and Andrienko [2], section 4.3). Thus, numeric data should be encoded by positions or sizes, while colour hues, shapes, and texture patterns are more suitable for qualitative data.

Ben Shneiderman ([41]) summarised the process of data exploration by means of visualisation in the well-known Information Seeking Mantra: "Overview first, zoom and filter, and then details-on-demand". For supporting an overall view of a dataset, it is necessary to visualise the data so that all visual elements representing data items could be perceived together as a single image (Bertin [6]). For further data exploration, visual displays need to be complemented with interactive tools for zooming, filtering, and accessing various details, or "drilling down" into the data (e.g. Buja et al. [7]).

All interactive tools need to be carefully designed for maximum user convenience and effectiveness of the exploration process. Direct manipulation methods, when the user interacts directly with a visual display, are highly recommendable. Mouse-operated widgets such as sliders and switches are also appropriate. Response time may be a critical issue in implementing interactive tools. It is desirable that the computer responds to an interactive operation within 50ms or at most 100ms; the user perceives this as an instantaneous response. However, in a case of a very large dataset, reaching such responsiveness may be extremely problematic.

When data have a complex structure (as, in particular, movement data, which involve space, time, population of entities, and a number of numeric and qualitative characteristics), they cannot be adequately visualised in a single display. Therefore, the use of multiple displays providing different perspectives into the data is important. The displays should be linked so that the information contained in individual views can be integrated into a coherent image of the data as a whole (Buja et al. [7]). The most popular method for linking parallel views is identical marking of corresponding parts of multiple displays, e.g. with the same colour or some other form of highlighting. Usually highlighting is applied to objects interactively selected by the user in one of the displays. This method, usually called "brushing", is a generalisation of the "scatterplot brushing" technique first implemented by Newton [38] and later elaborated in various directions.

The idea of brushing is illustrated in Fig. 1. Five displays show different aspects of the same data about movements of white storks in the course of their seasonal migration during 8 seasons from 1998 to 2006. The bar chart in the upper left corner represents the distribution of the movements by months of a year. The user has clicked on the highest bar, which corresponds to March. As a result, the data records about the movements that occurred in March have received a special status of selected records. All displays have reacted to this by marking the graphical elements corresponding to the selected records, as in the map and space-time cube in the lower part of the figure, or showing the positions and proportions of the selected data with regard to the whole dataset, as in the two histograms on the top.

Current approaches to display linking are described, e.g., by North and Shneiderman [39], Roberts [40], Baldonado and Woodruff [5].

Animated displays are often considered as the first choice when data involve time (Eick [14]). However, psychological studies show that animation is not necessarily effective and superior to static displays (Tversky et al. [48]). It seems that animation is good for gaining an initial overview of a time-related phenomenon or process while the further, more comprehensive exploration requires combination of animation with other displays and rich facilities for user interaction.

2.2 Visualisation of individual movement data

The early visualisations of movements on maps or in space-time cubes were produced manually, which was a laborious and time-consuming process. Computers and graphical display facilities not only simplified and expedited the work but also provided new opportunities, in particular, dynamics and possibility of user interaction with a display. Nowadays, animated maps ([4] [3]) and interactive cubes ([29] [25]) are widely used to visualise movement data.

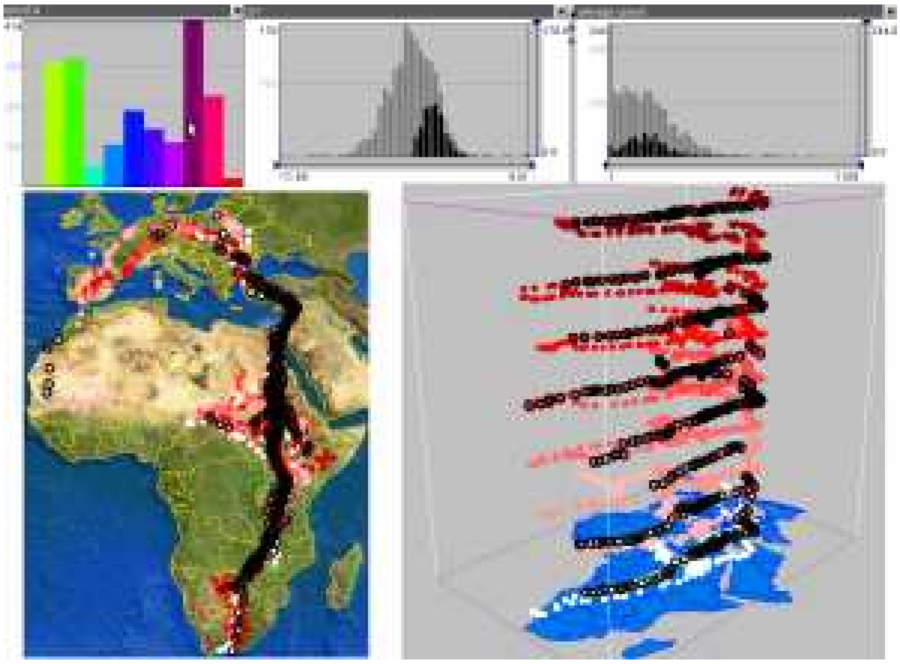

Fig. 1. An illustration of the technique of brushing between several parallel views of the same data. The illustration has been produced using the system CommonGIS. The satellite telemetry data about the seasonal migration of white storks have been collected in Vogelwarte Radolfzell, a department of the Max Planck Institute for Ornithology, Germany.

Map and cube displays are complemented with graphs and diagrams exhibiting various aspects of the movement ([29] [24] [37] [36] [13] [35] [34]). One example is time-time plot, or T-T-plot [24], which has two time axes and represents changes of a certain characteristic of the movement, such as the speed, travelled distance, or direction, between moments t_x and t_y by symbols placed in the positions corresponding to x and y or by colouring or shading of the cells, in which the plot area is divided.

A comprehensive research on methods for exploration of individual movement behaviours has been conducted in London City University ([37] [36] [13] [35] [34]). A specific focus of the researchers is very long trajectories, which require the use of data aggregation. Temporal aggregation occurs in a temporal histogram, which shows the number of visited locations by time intervals. Spatial aggregation is done by imposing a regular grid over the territory and counting trajectory points fitting in each cell. The resulting densities are visually represented by colouring or shading the grid cells on a map display. Densities counted for consecutive time intervals can be shown on an animated map display. A grid with densities can be treated as a surface, which may contain various features such as peaks (maxima), pits (minima), channels (linear minima), ridges (linear maxima), and saddles (channels crossing ridges). There are computational methods for detecting such features, which can then be visualised on a map.

In addition to the density surface, surfaces representing other movement-related characteristics may be built as suggested by Mountain [34]. Thus, an isochrone surface is a series of concentric polygons, centred on a selected location, representing the areas accessible from this location within specified "time budgets", e.g. 3, 6, 9 minutes, and so on. An accessibility surface is a grid where each cell represents the travel time from the selected location.

Laube et al. [32] represent movement behaviours of several entities such as football players in a matrix where the columns correspond to time intervals and the rows to the moving entities. Symbols or colouring in the cells of the matrix encode average characteristics of the movement of the entities, such as the speed or direction, on each of the intervals. Similarity of rows in such a matrix indicates that the respective entities have similar movement behaviours. The matrix is good for detecting certain types of patterns of collective movement, for example, "trend setting", when a group of entities repeats the movements of one entity after some time lag. However, with increasing the number of entities and the duration of the movement, the visual search for patterns becomes more and more difficult. It should be noted that visualisation of movement data has not been the main research focus of Laube et al. The researchers predominantly work on computational methods for automated detection of specific types of patterns in collective motion of groups of entities.

2.3 Visualisation of movements of multiple entities

Analogously to long time series of movement data, the visualisation of movements of numerous entities requires data aggregation or other ways of summarisation. Buliung and Kanaroglou [8] describe an approach where computational methods available in ArcGIS are applied to multiple trajectories. First, a convex hull containing all the trajectories is built. Then, the central tendency and the dispersion of the paths are computed and represented on a map. This method, however, is only suitable when the trajectories are sufficiently close to each other. When multiple entities synchronously move in the same direction, the visualisation described by Wilkinson [49] may be appropriate where the northerly migration of Monarch butterflies is shown on a map by "front lines" corresponding to different times. Again, this is a very special case.

Forer and Huismann [17] aggregate movement data into a surface by computing the total number of person-minutes spent in each cell of a regular grid. In a similar way, many other characteristics of multiple movements may be summarised and visualised. Kwan and Lee [31] build surfaces of summary characteristics of movements not in the geographical space but in an abstract space where the dimensions are the time of day and the distance from home. For this purpose, they use kernel density estimation methods. Such surfaces can be built for different groups of entities in order to compare their behaviours. Pairwise differences between surfaces can be computed and visualised.

Unfortunately, summarisation of movement data into surfaces severely alters their nature so that one can no longer see the changes of the spatial positions of the entities, i.e. the very essence of movement. To preserve the information about changes of positions, the data need to be aggregated in a different way. A possible approach is to count for each pair of locations (points or areas) in space how many entities moved from the first to the second location between two time moments. The resulting counts may be visualised as a transition matrix where the rows and columns correspond to the locations and symbols in the cells or cell colouring or shading encode the counts [20]. For more than one pairs of time moments, one would need to build several transition matrices, which could be then compared. However, the limitations of this approach with respect to the length of the time series of movement data are evident. Another problem is that such a visualisation lacks the spatial context. Some part of the spatial information may be preserved through ordering the spatial locations in the matrix in such a way that locations close in space are also close in the ordering. Guo and Gahegan [21] have made a survey of the existing methods applicable for this purpose.

Tobler [46] [47] suggests that numbers of entities or volumes of materials that moved from one place to another can be visualised by means of either discrete or continuous flow maps. A discrete map represents the movements by bands or arrows whose width is proportional to the volume moved (Fig. 2). Omitting minor flows increases map legibility when the number of locations is large. Continuous flow maps use vector fields or streamlines to show continuous flow patters (Fig. 3). According to Tobler, in a vector field the structure is immediately obvious, adjacent vectors clearly being correlated in length and direction. Conversely, if this is not the case then it is also obvious. Continuous flow maps are, in principle, not limited with regard to the number of different locations present in the original data. However, producing such maps from discrete data is computationally intensive.

Tobler's flow maps do not reflect the temporal dimension of movement data but show cumulative movements that occurred during a certain time period. However, the concept can be extended to animated flow maps or to series of flow maps showing how the flows change over time.

Cartographers Drecki and Forer [12] have designed a very interesting visualisation of aggregated movement data, specifically, movements of tourists coming to New Zealand (Fig. 4). They first transformed the travel times of the tourists from the absolute time scale (calendar dates) to a relative one starting from the day of tourist's arrival to New Zealand. Then the cartographers built a diagram consisting of six parallel planes, shown in a perspective view with a map of New Zealand depicted on each plane. The planes, from top to bottom,

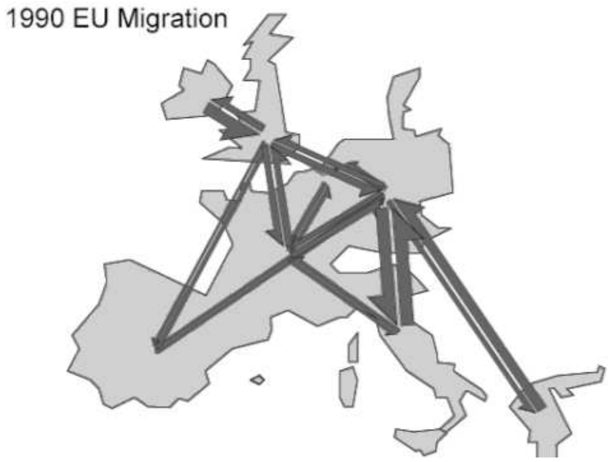

Fig. 2. A discrete flow map (Tobler [46] [47])

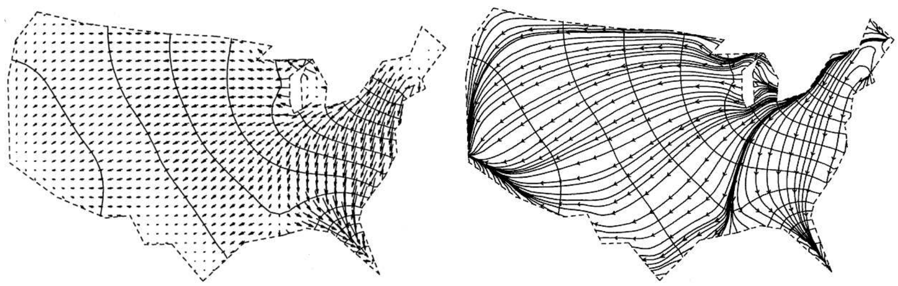

Fig. 3. Continuous flow maps (Tobler [46] [47])

correspond to the first six days of the tourists' travel through New Zealand. The movements of the tourists are shown by lines connecting the locations of the major tourist destinations on successive planes. The brightness of a line corresponds to the number of people that moved from its origin location (on the upper plane) to the destination location (on the lower plane) between the days corresponding to the upper and lower planes. To make the view clearer, the authors have omitted minor flows. The visualisation was designed for printing on paper. To our best knowledge, there are yet no software tools providing this kind of display for exploration of arbitrary movement data.

While software tools for data visualisation are usually supplied with appropriate user interaction facilities for view manipulation (e.g. zooming or rotation), filtering, querying display elements, selection, etc., there are approaches supposing that the user interacts with data prior to the visualisation. The idea is that the user selects a data subset of interest from a (possibly, very large) database, and only this subset is visualised. It is assumed that the size of the subset permits its visualisation without the use of aggregation. Kapler

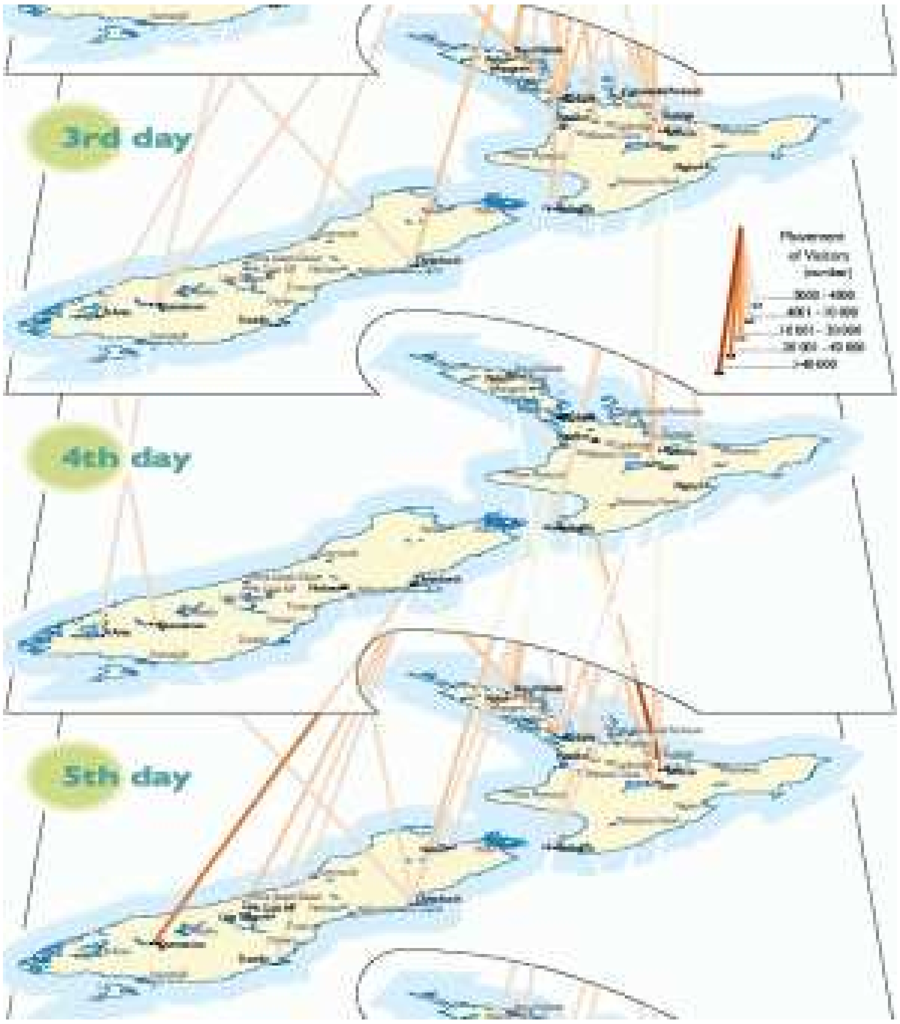

Fig. 4. A fragment of the visualisation of the movements of tourists in New Zealand created by Drecki and Forer [12]

and Wright [25] and Yu [50] apply this approach to movement data. Kapler and Wright suggest an ontology of movement data to support querying and search for information in the database. In particular, the user may consider the information on different levels of detail. In the system described by Yu, the user may formulate queries by referring to entities, their activities, and spatio-temporal relationships, specifically, co-location in space, co-location in time, and co-existence, i.e. co-location in both space and time.

It should be noted that the approaches based on selection and visualisation of small data subsets do not support an overall view of the collective behaviour of all entities and, hence, are insufficient for visual analysis of movement data.

2.4 Challenges

From the survey of the state of the art it may be seen that the research on methods for visual analysis of movement data did not reach yet its maturity. Most of the techniques and tools are not suitable for analysing data about many entities moving during long time periods. The limitations, which are recognised by many researchers ([31] [33]), come both from the side of the hardware (much computation time required, low speed of rendering, insufficient display size and resolution, etc.) and from the side of the user (display illegibility, perceptual and cognitive overload, difficulties in interpretation of unfamiliar visualisations and in operating complex visualisation environments). Hence, further research is required for finding ways to overcome these limitations.

Another problem is that each technique or tool allows one to consider movement data from a particular angle while the data are multifaceted and influenced by numerous factors, from characteristics of the moving entities to properties of the environment and various phenomena and processes occurring in it (see Chapter 1 in this volume). For a comprehensive analysis, several tools need to be combined. The selection of appropriate tools and methods should be based on a careful consideration of the needs of potential users (i.e. analysts of movement data) as well as their capabilities and limitations.

In the following sections, we try to apply a systematic approach to the selection and design of visual analytics methods for movement data based on the consideration of possible analysis questions the users may have, on the one hand, and the established principles of visual presentation of information, on the other hand (coming from the best practices in visual representation of information, these principles take into account, explicitly or implicitly, the perceptual and cognitive capabilities of humans).

3 Patterns in movement data

Chapter 1 in this volume defines the types of possible analytical questions about moving entities and stresses the primacy of synoptic questions, which involve multiple time moments and/or multiple entities considered all together. It introduces a generic concept of behaviour, the meaning of which embraces such notions as the trajectory of a single entity over a time period, the distribution of multiple entities in space at some time moment, and the collective movement of multiple entities in space over a time period. The primary objective in analysing movement data is to understand and characterise the movement behaviour of the entire population of moving entities over the whole time period the data refer to. On this basis, one can pursue further goals such as prediction of the future behaviour or optimisation of the movement.

Visual analytics mostly addresses the stage of gaining understanding and characterisation of behaviours. The objective of visual analytics in application to movement data may be stated as follows:

Allow a human analyst (also referred to as "the user") to understand and characterise the movement behaviour of a population of entities with the help of interactive visual displays, which are properly combined with other kinds of tools for analysis.

"Understand and characterise" a behaviour means represent it by an appropriate *pattern*. A pattern may be viewed as a statement in some language [16]. The language may be chosen quite arbitrarily (e.g. natural language, mathematical formulas, graphical language); hence, the syntactic and morphological features of a pattern are irrelevant to data analysis. What is relevant is the meaning, or semantics. It is natural to assume that representations of the same behaviour in different languages have a common meaning. Hence, the constructs of the different languages refer to the same system of basic language-independent elements from which various meanings can be composed. By analogy with meanings of words in a natural language, we can posit that the basic semantic elements for building various patterns include *pattern types* and *pattern properties*. A specific pattern is an *instantiation* of one or more pattern types. This is analogous to the specialisation of a general notion by means of appropriate qualifiers. In the case of patterns, the qualifiers are specific values of the pattern properties. For example, the pattern "entities e_1, e_2, ..., e_n moved together during the time period $\mathbf{T}$" instantiates the pattern type "joint movement" by specifying what entities and when moved in this manner.

It is quite reasonable to assume that the possible pattern types exist in the mind of a data analyst as mental schemata. Moreover, these schemata are likely to drive the process of visual data analysis, which is generally believed to be based on pattern recognition: the analyst looks for constructs that may be viewed as instantiations of the known pattern types. Therefore, for the design of proper visual analytics methods for movement data, it is important to define the pattern types relevant to such data.

3.1 Generic pattern types

On a very general level, pattern types are introduced in the book by Andrienko and Andrienko [2]. Descriptive patterns, which characterise behaviours, are distinguished from connectional patterns, which characterise relations between phenomena (see Chapter 1). The basic types of descriptive patterns are similarity, difference, and arrangement, where the latter type embraces such concepts as trend, sequence, periodicity, symmetry, etc. From instances of the basic pattern types, compound patterns are built as is shown graphically in Fig. 5.

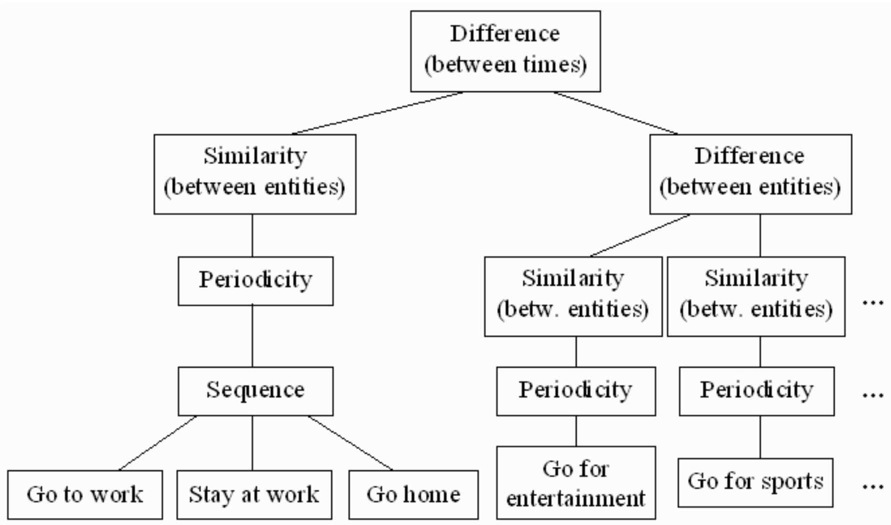

Fig. 5. An illustration of a compound pattern with several levels of nested sub-patterns

The types of connectional patterns are *correlation* (which is treated in a more general sense than just statistical correlation between numeric variables and includes co-occurrence of qualitative characteristics and co-occurrence of behavioural patterns), *influence* (or dependency, if viewed in the opposite direction), and *structure*, i.e. composition of a complex behaviour from simpler ones, like the visible movement of planets is a composition of their own movement and the movement of the Earth.

3.2 Descriptive pattern types for movement data

Let us now specialise these generic types of patterns for movement data. Our ultimate goal is to define pattern types for collective movement behaviours of multiple entities. In order to achieve this, it is necessary to consider the following "slices", or "projections", of this overall behaviour:

- individual movement behaviour, i.e. movement of a single entity over time, and
- distribution of movement characteristics (position, speed, direction, etc.) over the set of entities at a single time moment. For the sake of brevity, we shall call it "momentary collective behaviour".

We shall use the following abbreviations: **IMB** for individual movement behaviour, **MCB** for momentary collective behaviour, and **DCB** for dynamic collective behaviour, i.e. behaviour of multiple entities during a time interval. A DCB can be viewed from two different perspectives:

- As a construct formed from the IMBs of all entities, i.e. the behaviour (variation) of the IMB over the set of entities;
- As a construct formed from the MCBs at all time moments, i.e. the behaviour (variation) of the MCB over time.

These two views may be called aspectual behaviours [2]. They are essentially different and need to be described in terms of different types of patterns.

The variation of the IMB over the set of entities can be described by means of similarity and difference patterns, i.e. as groups of entities having similar IMBs, which differ from the IMBs in other groups of entities. For example, the weekday movement patterns of working people may be considered as similar. At the same time, they differ from the movement patterns of housewives and pensioners. It may happen that some entities have quite peculiar IMBs, which differ from the IMBs of all other entities. For instance, the movement behaviour of a tourist in a town may differ from the behaviours of the town residents. Such peculiar IMBs are also described by means of difference patterns.

Arrangement patterns are usually not relevant to the behaviour of the IMB over the set of entities because this set has no natural ordering and no distances between the elements [2].

What does it mean that IMBs of several entities are similar? There are diverse possible meanings, and all of them may be relevant to analysis of movement data:

- Similarity of the overall characteristics: geometric shapes of the trajectories, travelled distances, durations, movement vectors, etc.
- Co-location in space, i.e. the trajectories of the entities consist of the same positions or have some positions in common:
  - ordered co-location: the common positions are attained in the same order;
  - order-irrelevant co-location: the common positions may be attained in different orders;
  - symmetry: the common positions are attained in the opposite orders.
- Synchronisation in time:
  - full synchronisation: similar changes of movement characteristics occur at the same times;
  - lagged synchronisation: changes of the movement characteristics of entity e_1 are similar to changes of the movement characteristics of entity e_0 but occur after a time delay Δt.
- Co-incidence in space and time:
  - full co-incidence: same positions are attained at the same times;
  - lagged co-incidence: entity e_1 attains the same positions as entity e_0 but after a time delay Δt.

It should be noted that two or more IMBs may be similar during one time interval and dissimilar during another interval. Similarity and difference patterns may thus be applied not only to whole IMBs but also to their parts.

Let us now consider the other aspectual behaviour, that is, the behaviour of the MCB over time. Mathematically, time is a continuous set where ordering and distances exist between the elements, i.e. time moments. Hence, besides similarity and difference patterns, arrangement patterns are relevant. An arrangement pattern describes changes in the MCB with respect to the ordering and distances between the corresponding time moments. Here are the pattern types for describing the behaviour of the MCB over time (we note in parentheses the basic pattern types that have been specialised):

- Constancy (similarity): the MCB was the same or changed insignificantly during a time interval. For example, massive traffic towards industrial areas is observed during a time interval from 7AM till 9AM.
- Change (difference): the MCB significantly changed from moment t_1 to moment t_2. For instance, the movement in an area around a stadium after the beginning of a football game differs quite much from what could be observed before. Another abrupt change happens when the game is over.
- Trend (arrangement): consistent changes of the MCB during a time interval. For example, the traffic in industrial areas tends to gradually decrease after 9AM and tends to increase after 3PM.
- Fluctuation (arrangement): irregular changes of the MCB during an interval. Thus, the collective behaviour of vehicle drivers on a busy highway may irregularly vary depending on emergence of obstacles such as a traffic accident or just a truck trying to overtake another truck.
- Pattern change or pattern difference (difference): the behaviour of the MCB during time interval T_1 differs from that during time interval T_2. The term "pattern change" applies when T_1 and T_2 are adjacent. For example, a decreasing traffic trend between 9AM and 11AM changes for constancy between 11AM and 3PM, which, in turn, changes for an increasing trend after 3PM. The term "pattern difference" applies to non-adjacent time intervals.
- Sequence (arrangement): patterns follow one another in a specific order, such as traffic increase - constant heavy traffic - traffic decrease - constant low traffic and so on.
- Repetition (similarity): occurrences of the same patterns or pattern sequences on different time intervals. Thus, the traffic pattern sequences mentioned above occur every weekday.
- Periodicity, or regular repetition (similarity and arrangement): occurrences of the same patterns or pattern sequences on regularly spaced time intervals, like the weekday traffic patterns.
- Symmetry (similarity and arrangement): opposite trends like increase and decrease of traffic intensity; pattern sequences where the same patterns are arranged in opposite orders, for example, heavy traffic in the morning

followed by low traffic at midday in industrial areas and low traffic in the morning followed by heavy traffic at midday in touristic and shopping areas.

These pattern types are relevant not only to describing the variation of the MCB of the entire population but also to characterising movements of population subgroups. Thus, it seems reasonable to describe separately the variation of the MCB of car drivers and that of pedestrians.

3.3 Connectional patterns

The types of *correlation* and *influence* patterns are similar since the relation of influence differs from the relation of correlation only in its being directed: from two related things, it is specified which one influences the other. Correlations or influences may exist

- between different movement characteristics, e.g. direction and speed;
- between movement characteristics and supplementary characteristics (see Chapter 1), which include characteristics of entities, characteristics of time moments, and characteristics of spatial locations;
- between individual or collective movement behaviours on different time intervals, e.g. after slow movement in a traffic jam, drivers tend to move faster than usual;
- between collective movement behaviours of different subsets of entities, e.g. different teams in a football game;
- between individual or collective movement behaviours and supplementary characteristics, e.g. properties of the surface;
- between individual or collective movement behaviours and behaviours of external phenomena like weather, various types of events, etc.

What concerns *structure* patterns, these may be compositions of movement behaviours with regard to different temporal cycles like daily, weekly, and annual. For example, working people go to and from their work every day except weekends and go shopping on Saturdays. On Sundays, they usually stay at home in winter time and go to the countryside in summer time. Another example is the composition of the overall DCB in a city from the movements of the traffic, pedestrians, and cyclists, where each of the components is influenced by the others.

3.4 Pattern properties

When a user detects, for example, a pattern of similarity of IMBs of multiple entities, he/she is interested to know how many entities have this common behavioural pattern. Likewise, when the user detects a pattern of synchronisation or a trend, he/she measures the duration of the time interval when the entities moved synchronously or the trend lasted. These properties of the patterns taken as examples may be generalised as *support base*, that is, the size of the reference set on which this pattern takes place. Hence, for a pattern describing movements of multiple entities, the support base is the size of the subset of entities (i.e. the number of entities), and for a pattern describing movement on a time interval, the support base is the size (length) of the interval. Logically, the support base of a pattern describing movements of multiple entities during a time interval includes both the number of entities and the length of the interval. Besides the absolute support base, an important property is the *relative support base*, i.e. the size of the reference subset where the behaviour corresponds to the pattern in relation to the size of the whole reference set.

Not only is the length of the time interval of a pattern interesting for a user but also the *temporal localisation*, i.e. the position of the interval on the time scale. Likewise, it may be interesting to know which particular entities behave according to a pattern, in addition to the number of such entities. However, this pattern property may not always be accessible either because of a very large number of entities or because of privacy constraints.

Besides these general properties, there are also more specialised properties, which are relevant either to particular types of patterns or to characteristics involved in pattern definition. We shall not try to list exhaustively all these specialised properties but instead give a few examples. Thus, a change pattern may be characterised in terms of the magnitude and direction of the change while a periodic pattern is characterised by the length of the period between the repetitions. An important property of a similarity pattern describing movement of multiple entities along the same route (i.e. visiting the same locations in space) is the spatial localisation, i.e. where in space this common behaviour takes place. The properties of a pattern describing some behaviour in terms of the speed of movement include summarised speed characteristics such as average and maximum speed.

4 Helping users to detect patterns: a roadmap

It may be argued that the attempt to systemise and formalise movement patterns is more relevant to the development of computational analysis methods such as data mining than to visual analytics, where the detection and description of patterns involves such subjective processes as human perception and cognition. There are two major objections to this argument. First, depending on the presentation of material for human perception and cognition, these subjective processes can be either facilitated or impeded (they can even be purposefully manipulated, but this topic is out of the scope of our work). To facilitate the detection of patterns by appropriate presentation of data, we need to understand what types of patterns may exist in the data. On this basis, we can try to find visualisation and interaction techniques increasing the probability of such patterns being noticed by a human viewer.

The second objection is that visual analytics is not solely data visualisation but a synergistic work of human and computer supported by a synergy of visual and computational methods for data analysis. For achieving this synergy, it may be insufficient just to supply an analyst with independently developed visual and computational tools. It is more appropriate to design hybrid visuo-computational analysis methods and build corresponding tools. Knowing the types of patterns is essential for the design of such methods and tools. In particular, this may help to distribute the total analysis workload in an optimal way between the human and computer so that each side could apply its unique capabilities.

Let us give one example. A human analyst can effectively detect constancies, changes, trends and other patterns in the variation of the MCB over time by viewing animated map displays or map series. Computer methods, at least currently existing, can hardly surpass humans in grasping characteristic features of spatial distributions and their temporal dynamics. The role of computers is to help humans with preparation of the data and with testing of hypotheses gained. However, the variation of the IMBs over a large population of entities cannot be effectively investigated without involving computational techniques such as cluster analysis. One of the reasons is that none of the known visualisation techniques allows a legible representation of multiple IMBs. Another reason is that viewing of individual behaviours may be precluded for preserving the privacy of the individuals. Hence, the analyst needs to detect similarities and differences between IMBs as described in subsection 3.2 without seeing the IMBs themselves. The only possibility is to develop appropriate computational methods and techniques for representing their results.

According to the focus of this chapter, we shall not further discuss the computational techniques, which are sufficiently covered in the rest of the book. We shall mainly consider visualisation techniques as well as various data transformations, which may be required for effective visual exploration of movement data. In fact, the same or similar transformations may also be useful or necessary for preparing data to application of computational techniques. Moreover, it is important that visual and computational techniques are applied to the same data, either original or transformed, in order to extract complementary patterns contributing to comprehensive understanding of the data.

4.1 Data manipulation

Need for data aggregation

In designing methods and tools for helping users to recognise various patterns, we must comply with a crucial constraint: detecting patterns must be done without seeing any information about individual entities, for preserving their privacy. In other words, only aggregated or otherwise generalised data should be available to the user. Data aggregation may be indispensable also for another reason: the number of different entities and/or time moments may be so large that the visualisation of individual data becomes unfeasible because of the technical limitations (screen size and resolution) and/or impractical because of the human perceptual limitations. Hence, the role of data aggregation is both hiding individual information and reducing the amount of data.

While information reduction means substantial information loss, there is also a positive side, specifically, the possibility to omit "high-detail noise" and focus on characteristic features of the phenomenon under study. We may say that aggregation and generalisation helps us to see forest for trees.

The degree of data aggregation and generalisation matters a lot in data analysis. This is not only the matter of the size of the resulting data and the amount of information lost. This is also the matter of the scale on which the data are considered. Depending on the scale, the user sees the data differently and detects different patterns. Thus, in movement data, there may be local patterns like a flock (synchronous movement of several entities having close positions and same speed), or there may be larger scale patterns like massive movement towards industrial or commercial areas in mornings, or, on a yet larger scale, the difference of collective movement patterns on weekdays and weekends, and so on.

Hence, the appropriate degree of data aggregation and generalisation is not just a good trade-off between the simplification gained and the amount of information lost but it must be adequate to analysis goals. When the interests of the user include patterns of different scales, it is necessary to consider the data on different levels of aggregation. The tools for visual analysis must thus enable the user to do this, as is illustrated in Fig. 6.

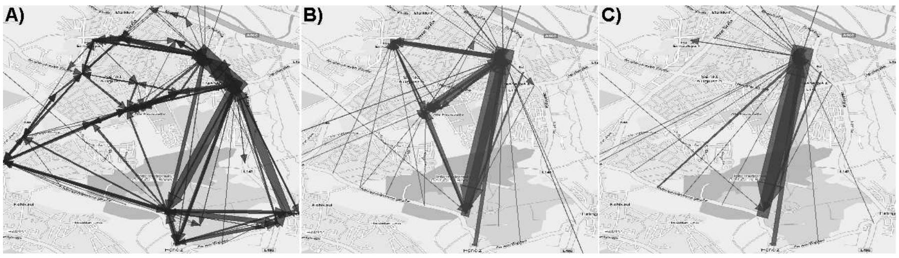

Fig. 6. Depending on the level of aggregation, different patterns can be observed in the same data about movements of a car. For the aggregation, the data have been divided into moves (trajectories) between stops of different duration: A) for 10 seconds or more; B) for 5 minutes or more; C) for 2 hours or more. Then, the numbers of moves between pairs of locations where the stops occurred have been counted and represented by vectors of proportional thickness. The illustration has been produced with the use of the system CommonGIS; the example data have been collected within the project GeoPKDD by GPS tracking of car positions.

Approaches to aggregation of movement data

Aggregation consists of two operations: 1) grouping of the individual data items or, in other words, division of the data into subsets and 2) deriving of characteristics of the subsets from the individual characteristics of their members. Typically, various statistical summaries are used as characteristics of the subsets: number of elements, mean, median, minimum, maximum values of characteristics, mode, percentiles, etc. It is also important to know the degree of variation of the characteristics within the aggregates. For this purposes, such statistical measures as variance (or standard deviation) or distance between the quartiles are computed. Aggregates with high variation of characteristics of the members should be avoided since they may lead to wrong conclusions concerning the data. Grouping/division may be necessary not only for data aggregation but also for other kinds of data processing and analysis.

Primary items in movement data are usually tuples (records) consisting of entity identifiers, references to time moments, references to positions in space and, possibly, values of movement characteristics such as speed and direction. These micro-items are typically combined into larger structures. The largest are the so-called *lifelines* of the entities where a lifeline includes all micro-items referring to the same entity. Lifelines are often divided into trajectories or into movement episodes. A *trajectory* is a sequence of items corresponding to a trip of an entity from one location (source) to another (destination) where the source and destination are defined semantically (e.g. home, work, shop, etc.) or according to the time the entity spends in a location. *Movement episodes* [13] are fragments of lifelines where the movement characteristics (speed, direction, sinuosity, etc.) are relatively constant whereas a significant change indicates the beginning of the next episode. Movement episodes, trajectories, and lifelines can be viewed as macro-items of movement data. An analytical toolkit should enable the user to unite micro-items into macro-items according to various criteria. Thus, in Fig. 6, micro-items have been combined into trajectories. The sources and destinations have been chosen according to the time spent in a location. Depending on the choice of the minimum time threshold, shorter or longer trajectories are obtained.

There are also other possible methods of grouping, which can be applied either to micro-items or to macro-items. Thus, micro-items or short trajectories or movement episodes may be grouped according to the time of their occurrence. For this purpose, the whole time span is divided into intervals, and items occurring on the same intervals are grouped together, as in Fig.9. Depending on the data and analysis goals, it may be useful to divide the time into equal intervals (e.g. 10 minutes, 1 hour, or 1 week), or into slightly unequal intervals corresponding to calendar units such as months, quarters, or years, or to apply other division principles, for example, divide a year into semesters and holidays. Furthermore, it may be reasonable to divide the time into subsets consisting of non-contiguous intervals, in particular, according to one or more of the temporal cycles. Thus, the user may wish to group all Mondays, all Tuesdays, and so on. Hence, the data analytics toolkit should include a tool for time partitioning where the user can flexibly define the principles of division.

Another grouping principle applicable to micro-items is according to the places where they occur. For this purpose, the space is divided into compartments, which may be cells of a regular grid, units of administrative or other existing territory division, or areas defined by the user according to any appropriate criteria such as surface type, way of use, accessibility, or other relevant properties of the space. The visual analytics tools should support such arbitrary divisions of the space. Thus, the user may define space compartments by interacting with a map display or by applying database search and computational operations like retrieving the locations of schools, shops, etc. and building buffer zones around them. Generally, it is not necessary that user-defined space compartments cover the whole territory since there may be places never visited by the moving entities under analysis.

Space-based grouping is also applicable to movement episodes and trajectories, but in a different way: macro-items are grouped together if they start and/or end in the same compartments. This method of grouping has been applied in Figs. 6 and 9. The space compartments have been defined as circles encompassing the locations where the entities stopped.

It is also possible to group micro- and macro-items according to values of various attributes including movement characteristics (speed, direction, transportation means, etc.) and characteristics of the entities (e.g. age or occupation in case of people). Since movement characteristics in macro-items are not constant but change over time, grouping can be done on the basis of values at selected time moments or on the basis of aggregated values on selected time intervals. Unfortunately, selection of each additional time moment or interval multiplies the number of groups and causes difficulties for the visualisation and visual exploration of the results of the aggregation. Besides attribute values at selected time moments, macro-items can also be grouped on the basis of *changes* of the values that occurred between two time moments.

Other data transformations

Aggregation is not the only useful data transformation, and we shall briefly discuss some other data manipulation techniques that may increase the comprehensiveness of analysis and give additional insights into the data. One of them is the computation of the amounts or degrees and directions of changes, which is valuable not only for grouping of the entities by also by itself. Thus, it may be useful to look at change maps portraying (in a generalised manner) the changes of the MCB from one moment to another.

From other possible methods, especially useful may be transformations of space and time from absolute to relative. Thus, similarities between temporally and/or spatially separated behaviours represented by lifelines or trajectories can be more easily detected when these behaviours are somehow aligned in time and/or in space. To align behaviours in time, the "objective", absolute time of each behaviour (i.e. the calendar dates and times) is ignored and only its "internal" time is considered, i.e. the time relative to the moment when this behaviour began. Thus, in the representation of the tourist movement in New Zealand (Fig. 4), the analysts superposed the starting times of the IMBs of different tourists. It may also be useful to superpose both starting and ending times. In this case, the absolute time moments in each IMB are transformed into their distances from the starting moment divided by the durations of the behaviours (i.e. the lengths of the intervals between the starting and ending moments). This facilitates detecting similarities between movements performed with different speeds. Such an approach could be useful, for example, in comparing movements of cars and bycicles.

Analogous ideas can be applied for spatial alignment of trajectories or lifelines initially disjoint in space. A user may try to bring a set of trajectories to a common origin and search for coincidences between them. Furthermore, the user may be interested in disregarding the movement directions and considering only changes of the direction (turns). For this purpose, the trajectories are "rotated" until the initial movement directions coincide. Coincidences between further trajectory fragments indicate similarities. It may also be useful to "stretch" or "shrink" the trajectories to adjust their lengths.

In looking for co-locations between trajectories where positions are specified as points in the space, it may be reasonable to apply a kind of "spatial coarsening", i.e. replace the original points by regions (areas), for example, circles with some chosen radius around the points. The resulting trajectories are treated as similar when there is an overlap between their "expanded" positions while there may be no sharp co-incidence between the original positions.

In studying MCBs and their behaviours over time, it may be appropriate to treat the space as a discrete set of coarsely defined "places" rather as a continuous set consisting of dimensionless points. For this purpose, one uses the methods for space partitioning, which has been discussed before in relation to data aggregation. Such a transformation may be called "space discretisation". Furthermore, it may be useful to transform the geographical space into a kind of "semantic" space consisting of such locations as home, working place, shopping site, sport facility, etc. Then, each trajectory is transformed into a sequence of movements between pairs of these locations, and the user looks for similar sub-sequences occurring in different trajectories.

4.2 Visualisation and interaction

As we have mentioned earlier, a DCB of a set of entities over a period of time involves two aspects, the variation of the IMB over the set of entities and the variation of the MCB over the time. Different types of patterns are relevant to each aspect and, hence, different tools are needed to support the detection of pattern instances.

The pattern types corresponding to the first aspect are patterns of similarity and difference between IMBs. As we have already noted, the challenge is that similarities and differences have to be detected without the user seeing the IMBs, for the reasons of privacy and data size. The only possible solution is computational search for similarities and differences, for example, with the use of clustering methods. It would be reasonable to develop a clustering tool that allows the user to specify the kind of similarity he/she is currently interested in (see Section 3.2) and uses an appropriate function for computing the degrees of similarity from a library of possible functions. The results of clustering need to be visualised so that the user could interpret and investigate them. A general approach is to display various statistics about each of the clusters, i.e. aggregated data obtained from individual characteristics of the members of the clusters.

The exploration of the second aspect of the DCB may be supported by visual displays that represent the MCB at different times. There are two generic ways to do this: display animation and display iteration. In display animation, the views of the MCB at different times are arranged temporally and presented one by one. In display iteration, these views are arranged spatially (within the space of the screen) and presented simultaneously. Animation or iteration may be applied to various types of displays, such as maps, diagrams, or graphs. In the context of this study, it is essential to use displays representing aggregated rather than individual data, for the reasons of data size and privacy.

Hence, the exploration and analysis of both aspects of the dynamic collective behaviour require the visualisation of aggregated movement data. The difference is only in the way the data are aggregated, through clustering or through interactive division. Therefore, the same visualisation techniques may be applicable in both cases. Let us now review the methods suitable for the visualisation of aggregated movement data.

Maps and map-based displays

Map displays are best suited for the visualisation of spatial data such as positions, directions, and trajectories as well as non-spatial data associated with positions, directions, and trajectories. In our case, the data must be displayed in an aggregated or generalised way. Thus, instead of the individual positions of entities, the densities of the entities at various places may be visualised. The densities may be computed from data referring to a single time moment or to a time interval. In the latter case, both spatial and temporal aggregation take place.

There are two principal approaches to displaying densities. One of them is to build a smooth density surface using appropriate computational methods for spatial interpolation between original positions (see, for example, [42], Chapter 14). In cartography, there are several methods for portraying such surfaces ([42], Chapter 15), in particular, contour lines, or isolines (projected intersections of the surface with horizontal planes corresponding to selected values), hypsometric tints (shaded areas between the contour lines), and continuous-tone map, in which each point of the surface is shaded with a grey tone or colour proportional to the value of the surface at that point. The surface can also be given a three-dimensional look in a perspective view.

Another approach is to compute aggregated values for areas, for example, cells of a regular grid, and represent them on a map as characteristics of the areas. This can be done, for example, in the way of shading or colouring the areas according to the respective values or drawing inside the areas symbols or diagrams with the sizes proportional to the values (Fig. 7).

While smoothing may be good for exposing large-scale patterns of MCB, the display of non-smoothed aggregated data may be equally effective for this purpose and at the same time serve better for detecting local peculiarities. Moreover, not only densities can be displayed in this way but also other aggregated movement characteristics, for example, average speeds or travel distances. Thus, the iterated maps in Fig. 7 show data about movements of storks that have been aggregated spatially by cells of a regular rectangular grid and temporally by months. The graduated circles in the cells represent the average speeds of the birds' movement within the cells during the corresponding months, from August 1999 (top left) to April 2000 (bottom right). It should be noted that smoothing would be hardly effective for data like these, where the movements are not spread over the whole territory. Hence, data aggregation on the basis of space discretisation has a more general applicability than summarisation of data into smooth surfaces.

To explore data about movement directions, the user may be suggested a map display showing the prevailing movement directions in different places, which may look like the vector map on the left of Fig. 3. A vector map may show not only the prevailing direction in each place (by vector orientation) but also how many entities moved in that direction (e.g. by vector length) and how much this direction prevails in relation to the other directions (e.g. by vector shade or colour).

However, it is not always the case when one direction significantly prevails over others. Therefore, it may be reasonable to look also at more detailed information, e.g. how many entities moved in each direction in any place over the territory. For a single place, this information may be portrayed by means of diagrams as shown in Fig. 8. In the diagrams B and C, the size of the internal circle may encode the number of entities that stayed in the same place. Multiple diagrams may be overlaid on a map to show movement directions in different places. To avoid overlapping, the diagrams have to fit into the corresponding space compartments, but if the compartments are very small, the diagrams may be illegible.

To see not only the movement directions but also how far the entities moved, one can apply a discrete flow map technique. Such a map, as any other, may be animated or iterated to represent movements done over a period of

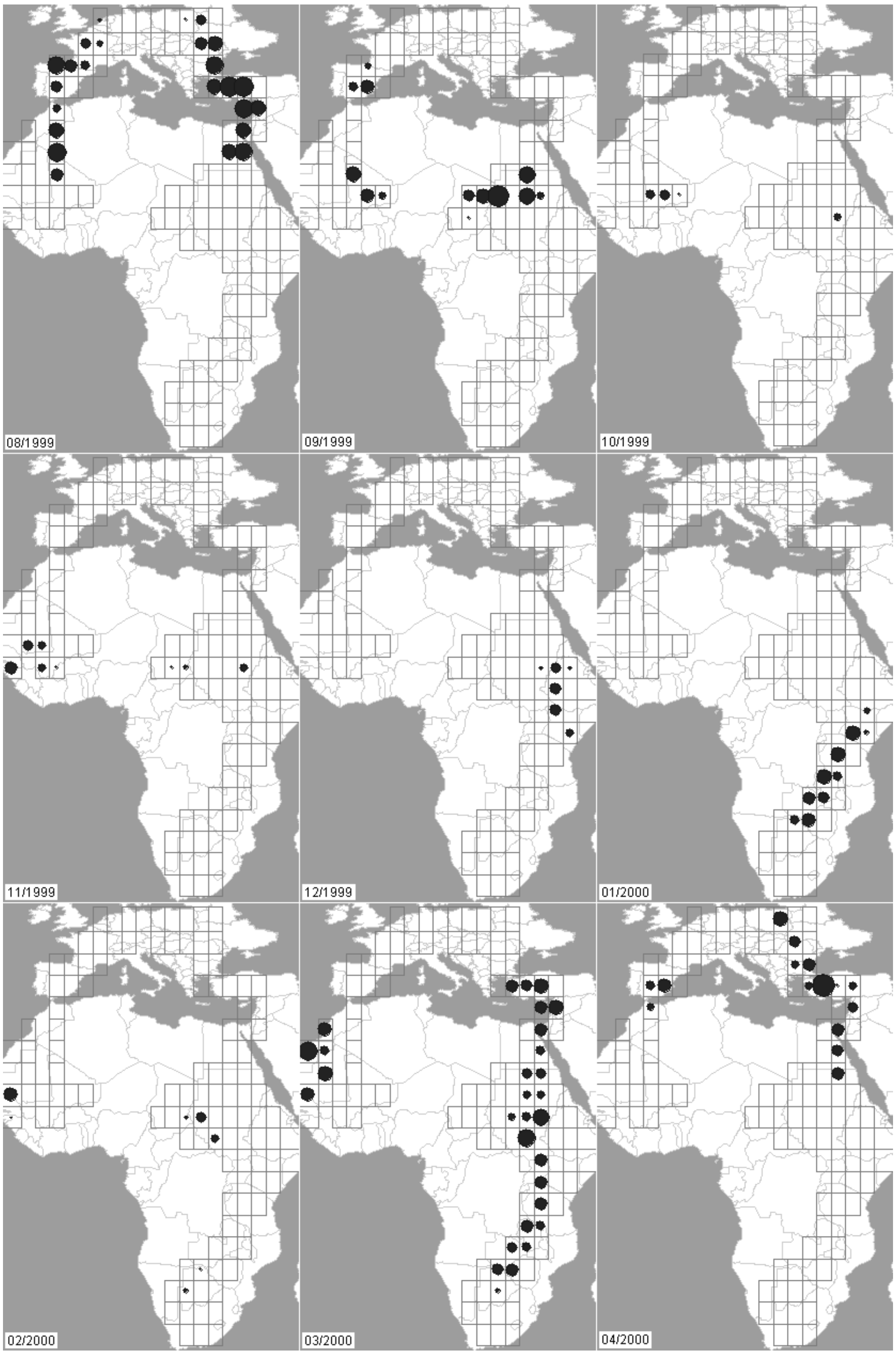

Fig. 7. A visualisation of the data about the movements of storks (introduced in Fig.1) aggregated spatially by cells of a regular rectangular grid and temporally by months.

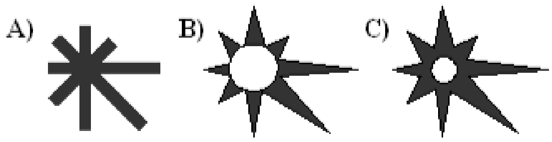

Fig. 8. Possible methods for representing numbers of entities moving in different directions.

time (Fig. 9). Another possibility is to use a three-dimensional view, such as the tiered maps in the visualisation of the tourist movement in New Zealand (Fig. 4).

In any variety of the discrete flow map technique, there is a risk that a large number of flows and intersections between them can make the display illegible. Additionally, iterated or tiered maps have limitations in the number of consecutive time moments that can be shown. Interactive techniques can compensate for these limitations, at least partly. Thus, interactive filtering can remove minor flows from the display. This reduces overlapping and allows the user to focus on major flows. Another useful operation is filtering according to the source or destination location. The user can also be given controls for "temporal scrolling", i.e. shifting the time period reflected in the iterated map display or map tiers. This alleviates the limitation concerning the number of time moments.

Non-cartographic displays

As maps and map-based displays cannot adequately reflect all relevant aspects of complex data such as movement, they need to be complemented with other types of display. One of them is frequency histogram known from statistics. It shows the distribution of numeric or qualitative characteristics over some population, which may be, in particular, a population of moving entities (see Fig.1 top right). Histograms may be used for the exploration of the frequencies of different values of speed, direction, acceleration, etc. at selected time moments or over the whole time and for comparison of the distributions at different time moments. Histograms representing the frequencies of different movement directions may have polar or star-like layout rather than linear (see Fig. 8A). Histogram displays may allow interactive brushing as described by Spence and Tweedy [43].

There is an extension of the histogram technique known as two-dimensional histogram or binned scatterplot [11]. The plot has two axes corresponding to value ranges of two selected attributes. The area of the plot is divided into regular compartments (bins), in which the frequencies of the corresponding value combinations are shown by symbol sizes, shading, or colouring. Analogously, other aggregated characteristics may be represented, for instance, average or median values of another attribute. In the exploration of time-referenced data, it may be useful to apply a binned scatterplot where the axes correspond to

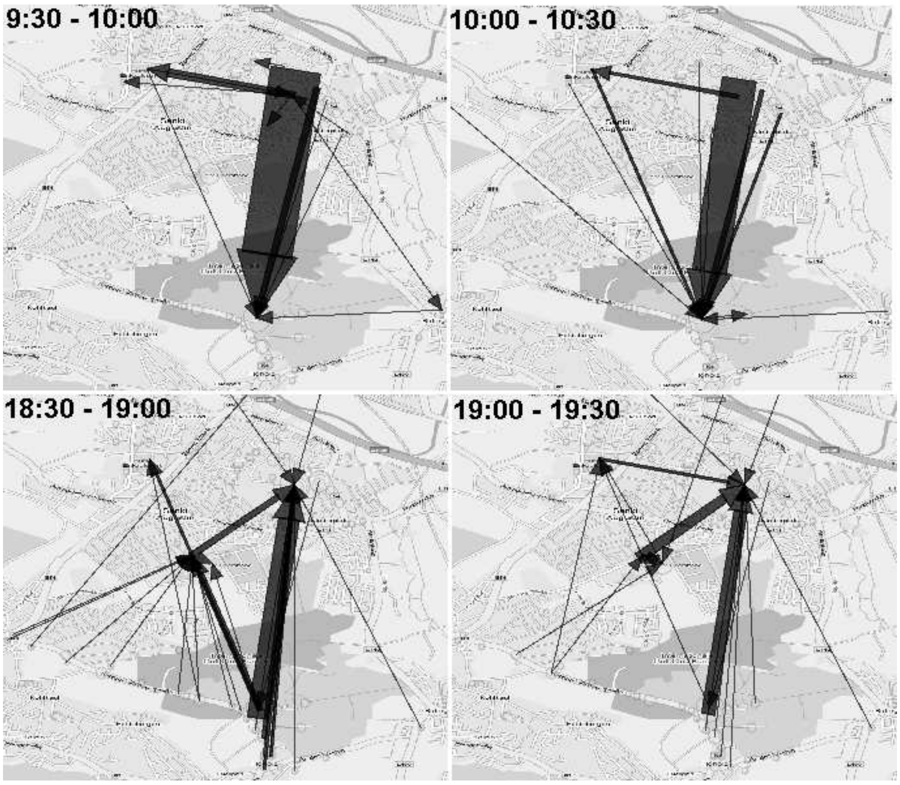

Fig. 9. An illustration of the technique of animated or iterated discrete flow map. The same data as in Fig. 6 have been divided into trajectories using the locations of the stops at least 5 minutes long as the sources and destinations. The space has been discretised by building circles around the source and destination locations. The time of the day has been divided into 30-minute intervals. Each map in the series or each frame in the animation represents the movements having occurred during the corresponding intervals. Specifically, the widths of the arrows connecting the coarsened sources and destinations are proportional to the number of entities that moved between these locations.

temporal cycles. Thus, in Fig. 10, the horizontal axes of the two plots correspond to the time of a day divided into 1-hour intervals while the vertical axes correspond to the days of a week, from 1 (Monday) to 7 (Sunday).

A variation of two-dimensional histogram is a transition matrix (Fig. 11), where rows and columns correspond to different spatial locations while the symbol or colour in each cell shows how many entities moved from one of the respective locations to the other between the selected time moments.

Fig. 12 demonstrates how aggregated data may be visualised by means of segmented bars. One of the dimensions of the diagram may represent time or the value range of any selected attribute. The other dimension represents the

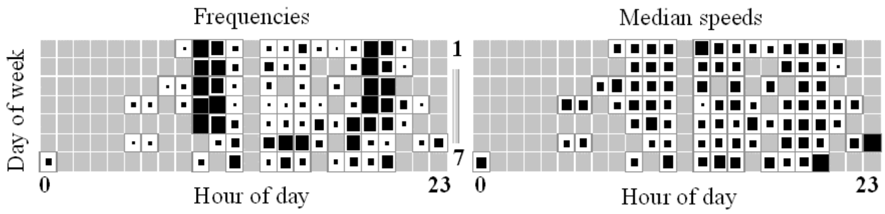

Fig. 10. Illustration of the technique of binned scatterplot. The car movement data introduced in Fig. 6 have been aggregated by days of week and hours of day. It may be noted that no movements occurred from hour 1 till 5 and from 13 till 14.

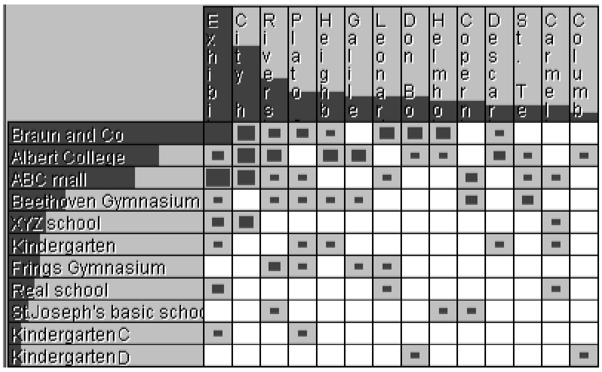

Fig. 11. A transition matrix built from simulated data about transportation of people. The display has been produced using the system CommonGIS.

numbers of corresponding data items or entities divided into subsets according to values of another attribute.

As mentioned in section 2.2, a useful technique supporting the exploration of time-related data is T-T-plot [24]. However, T-T plots represent information about individual entities, which is inappropriate in our case. Therefore, the technique needs to be modified to show aggregated information. Instead of the changes of the individual values, the plot may show statistical summaries of the changes over the whole population or a group of entities, such as the means or medians.

Multiple linked views

When data have complex structure, there is no way to represent all the information in a single display. As may be seen from the discussion of various visualisation techniques suitable for movement data, each of them shows only a certain aspect of the data. Therefore, different techniques need to be combined for a comprehensive visual exploration of the data. Moreover, these different techniques need to be used in parallel; otherwise, the user will not be

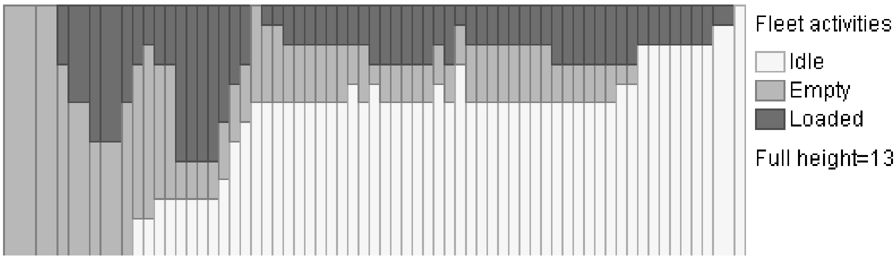

Fig. 12. Illustration of the technique of segmented bars. Each bar corresponds to a time interval. The full height of a bar represents the number of moving entities (in this case vehicles transporting people or goods) and divided into segments according to certain movement characteristics (in this case activity type: absence of motion, movement without load, or movement with load). The time span has been divided into intervals according to moments when relevant changes occurred in the situation. For this reason, the bars differ in their widths. The display has been built in the system CommonGIS from simulated transportation data.

able to relate, for example, the distribution of the entities in space at a particular moment to changes of their positions and other movement characteristics. The user should be provided interactive facilities to support establishing connections between different views, in particular, finding elements in different displays corresponding to same spatial positions, same times, same groups of entities, and/or same values of movement characteristics.

Brushing has been mentioned several times as a technique that supports establishing links between displays. Besides brushing, there are other methods of display linking such as propagation of a division of the set of entities into classes (each class is assigned its specific colour, which is consistently used in all displays) and simultaneous reaction of all displays to interactive filtering of the data; a review may be found in [2]. It should be noted that most of the currently existing techniques for coordination of multiple displays involve dealing with individual data items and are therefore not scalable to very large datasets. There is a need in new technical solutions, which could properly work in a situation when all displays show only aggregated data (moreover, differently aggregated data!) while the original individual data are not present in the computer memory.

4.3 Supporting search for connectional patterns

Linking of two or more displays providing complementary information may be helpful in a search for connectional patterns, i.e. for correlations, influences, and structural links between characteristics, phenomena, processes, events, etc. For example, linking can help the user to relate speeds represented on a histogram to positions on a map. Another approach is to represent the different phenomena or characteristics on the same display. The most common techniques suitable for this purpose are (binned) scatterplots and maps. Scatterplots expose correlations between values of two numeric characteristics. Maps combine representations of two or more characteristics or phenomena by overlaying several information layers. For example, movement data may be represented as flows or vectors on top of a layer representing weather or land cover information by area painting and a layer representing rivers and other waters by lines or shapes of a specific colour.

For other display types, ways to incorporate additional information can sometimes be found while there are no general approaches. Thus, in a time series display like in Fig. 12, moments of various events may be represented as ticks on the temporal axis, which may be differently coloured to indicate event types.

In order to investigate movement data for structure patterns related to various temporal cycles, it may be helpful to look at iterated displays arranged according to the cycles. For example, maps of city traffic aggregated by days may be arranged on the screen into a matrix with 7 columns corresponding to the days of the week and the rows corresponding to different weeks. This arrangement facilitates noticing commonalities and differences within and between the cycles.

The user may also benefit from a temporal query tool capable of extracting data that refer to the same relative positions or sub-intervals in different cycles and aggregating the data across the cycles. For example, the user may be interested to extract all people movements made from 6AM to 9AM in all days and have the extracted data aggregated by the days of the week. Then, the aggregated morning hours movements on Mondays, Tuesdays, Wednesdays, and so on should be appropriately presented to the user so that the user could see the differences between the movements on weekdays and weekends. Moreover, the movements on Monday mornings may differ from the movements in the mornings of other weekdays and movements on Saturday mornings may differ from those on Sunday mornings. Similar queries can be applied to other hours of the day in order to understand in the result how the daily and weekly cycles interact in people movement.

5 Visualisation of patterns

5.1 Need for pattern visualisation

We have discussed a number of visual and interactive techniques intended to help users in detecting patterns in movement data. According to the philosophy of visual analytics, visual search for patterns by a human can and should be complemented by automated search, or "mining", since computers may be able to discover such types of patterns that are hard to notice by a human, and vice versa. The use of automatic methods requires the results to be presented in a way allowing the user to interpret and evaluate them. In other words, automatically detected patterns need to be made perceptible to human mind. However, this requirement is also valid for patterns detected by visual methods: a person who has observed a pattern needs to represent it in such a way that it could be perceptible to other people as well as to this person after some time. Hence, irrespective of the method used for finding patterns, there is a need in their explicit *visualisation* (let us recall that the word "visualise" is defined in a dictionary as "to make perceptible to the mind or imagination").

To our knowledge, the research on the visualisation of patterns extracted from data is currently in its infancy and consists mainly of a few ad hoc methods devised for particular types of data mining results. In the area of data visualisation, the researchers are focused on the task of enabling users to detect patterns and do not consider the problem of how these patterns can be explicitly represented. There is a more general research on knowledge visualisation [44] conducted in the fields of knowledge management and education. Most methods suggested for knowledge visualisation are based on the use of node-link structures, or graphs. This includes semantic networks, also known as concept maps or cognitive maps, mind maps, argumentation maps, storyboards, etc. Generally, graphs are quite powerful instruments for representing various relationships and are therefore used for the visualisation of some types of data mining results such as association rules and decision trees.

There are yet no specific methods for the visualisation of patterns extracted from movement data. To find approaches to creating such methods, it seems reasonable to start with reviewing the existing methods used for pattern visualisation in the area of data mining and knowledge discovery in databases.

It should be noted that the ideas of visual analytics are quite similar to the concept of *visual data mining* [28] [26], which means involvement of visual representations in all stages of the data-mining life cycle, including data preparation, model derivation, and validation [18]. The goal is to achieve a synergy of visualisation and data mining and to enhance the effectiveness of the overall data mining process. Despite the similarity, visual analytics has a broader scope embracing visualisation, interaction, various data transformations (not only as preparation to data mining), computational analysis methods, support of analytical reasoning, collaborative deliberation, and visual communication.

Presented below is a very brief overview of the most common forms of data mining results and the approaches to their visualisation. For a more detailed survey, see [27]. Along with the descriptions of the existing approaches, we try to speculate how they can be adapted to the specific pattern types that may be discovered in movement data. This is not an easy task, first of all because of the necessity to deal with space and time, which substantially differ by their nature from numeric and nominal variables and therefore require special visualisation techniques.

5.2 Visualisation of clusters

Clustering algorithms group various kinds of objects according to similarity (closeness) of their characteristics, and the user needs to understand what the objects in each cluster have in common. Unfortunately, clustering algorithms do not provide any general description of the clusters built. The clusters are defined extensionally, i.e. by listing the elements they consist of. Hence, any information about the common features of the objects in each cluster has to be extracted from the data that were used as the input of the clustering method. A realistic way to do this is to obtain various statistics about the characteristics of the members of a cluster and to visualise these statistics. By comparing the statistics for different clusters, the user can understand what is in common between the members of each cluster and how they differ from the members of the other clusters.

The general approach to the visualisation of clustering results is illustrated in Fig. 13, which represents nine clusters, with their relative sizes (in percentage to the size of the whole dataset) shown by bar segments and numbers on the left of the picture. The pie charts and bar charts represent the distributions of categorical and numeric characteristics, respectively, within each cluster in comparison to the distributions of these characteristics in the entire dataset.

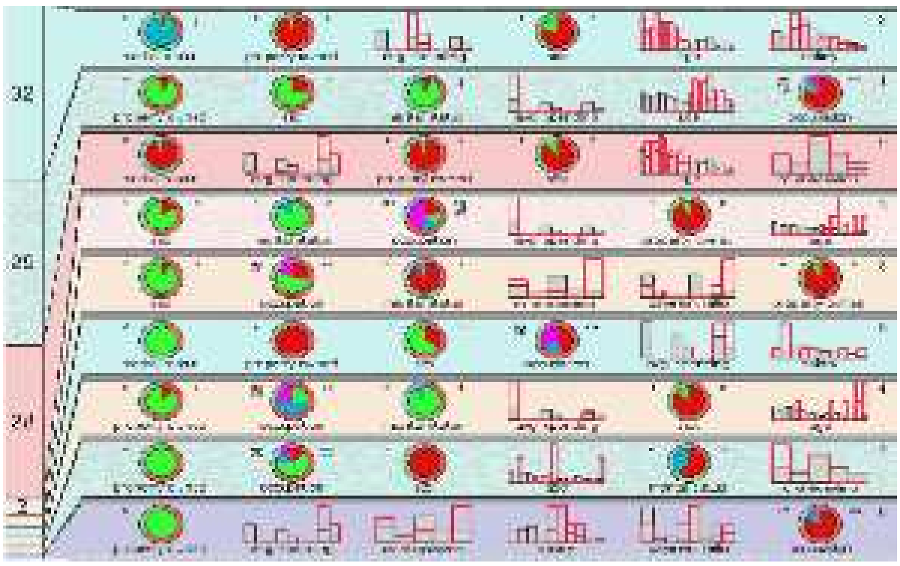

Fig. 13. Representation of clustering results in IBM Intelligent Miner.

As has been mentioned, clustering is an appropriate instrument for the analysis of the variation of individual movement behaviours (IMBs) and their parts (trajectories or movement episodes). A clustering method divides these macro-items into groups according to a selected definition of similarity (see section 3.2). The user then needs to see the common features of the items in each cluster as well as the degree of variation. As in the general case, this can be done by computing and visualising statistics about the items in the clusters. Appropriate statistics and visualisation techniques depend on the chosen definition of similarity.

Thus, when the similarity is defined as co-location of the trajectories, a suitable visualisation of a cluster would be a map showing for each location (resulting from space coarsening or original, if there are not too many different locations in the source data) in how many trajectories it appears. Graduated symbols or graduated shading are suitable for this purpose. A separate map is built for each cluster, which enables comparison of the clusters. For ordered co-location and for spatio-temporal co-incidence, it is reasonable to compute for each pair of locations x and y and time interval T how many cluster members moved from x to y during the interval T, where T results from an appropriate partitioning of the time (which may be previously transformed as discussed earlier). A possible way to visualise these statistics is an animated or iterated flow map, as in Fig. 9, or a three-dimensional map-based display (tiered maps), as in Fig. 4.

Unlike the pie charts and histograms in Fig. 13, map-based displays representing different clusters of trajectories are quite sizable and cannot be easily put together on a single screen. This means that the user will not be able to see the information about all clusters simultaneously and will need tools for browsing through the set of clusters and selecting pairs for comparison.

5.3 Visualisation of association rules

While there is extended research concerning association rules, the main focus of it is how to extract rules efficiently. Limited work has been done on how to enhance the comprehension of the discovered rules. The major problems that need to be addressed are the large number of association rules often generated, the difficulty of comprehending the output format that they have, and the difficulty of interpreting their specific semantic information [51]. Association rules that may be mined from movement data will pose additional problems, as indicated in Section 5.1.

At the present time, there are several approaches to the visualisation of association rules. Some of them, like VisualMine [22], focus primarily on relationships among the items occurring in the rules but do not provide any techniques for the user to explore relationships among different rules. Other tools (e.g. [23]) allow the user to analyse individual rules in detail while loosing the overview of the entire rule set. Buono [9] introduces a graph-based technique that allows the user to visualise a great number of association rules. Interaction tools allow the user to manipulate the graph and explore the rules.

Fig.14 gives an illustration of a possible visualisation of association rules, specifically, the Rules Graph technique from the commercially available data-mining package IBM DB2 Intelligent Miner for Data ($http : //www - 306.ibm.com/software/data/iminer/fordata/$). The nodes of the graph represent the item sets occurring in the bodies and heads of the rules and the arrows represent the rules. The colour of an arrow represents the importance measure of the corresponding rule (more precisely, the lift, which is defined as the confidence of the rule divided by the support of the rule head) and the width shows the confidence of the rule. Colours of the nodes represent the supports of the corresponding item sets.

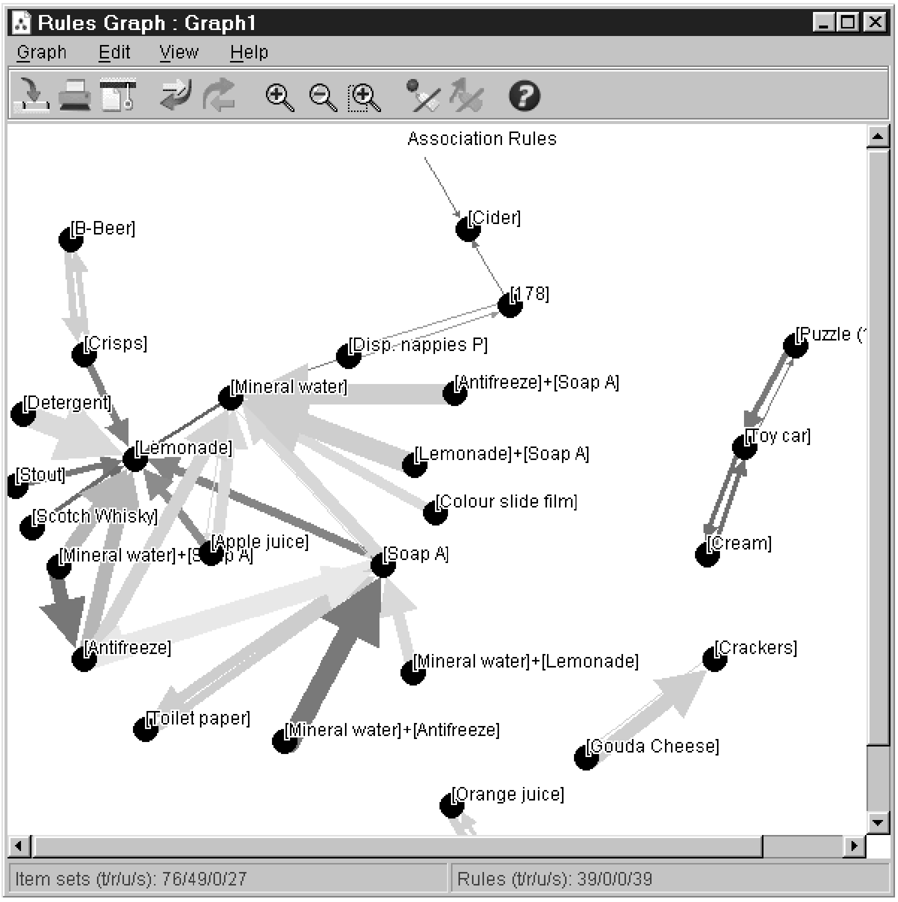

Fig. 14. A set of association rules visualised in IBM Intelligent Miner software by means of Rule Graph technique.

This visualisation demonstrates the use of node-link structures, which is, perhaps, the most fundamental and widely used method for knowledge representation [44]. Note that arrow symbols are paramount for visual communication; they are used multi-purposely to represent directions, movements, orders, relations, interactions, and so forth [30]. This makes node-link structures quite suitable for the visualisation of some types of patterns that can be extracted from movement data, in particular, temporally annotated sequential patterns [19], which may be related to locations or regions in space. An example of such a pattern is a frequently appearing sequence of places A, B, C with the transition time from A to B being t_1 and the transition time from B to C being t_2 (t_1 and t_2 may be average times or intervals). Nodes may be used to represent the places and links (arrows) may indicate the temporal order in which these places were visited. The arrow symbols may differ in width, colour, brightness, and/or texture to represent the transition times and other characteristics. Additionally, text labels may be used. The graph may be drawn on top of a map to allow the user to recognise the places, see their relative positions, and relate the patterns to various geographical information.

However, simultaneous visualisation of all sequential patterns extracted from movement data may be impracticable not only because of their potentially great number but also because of possible overlaps between the patterns when one and the same place appears in two or more patterns. Therefore, additional tools are required for navigation through the set of patterns and selection of patterns for more detailed examination and comparison. The user should be able to select subsets of the patterns according to various criteria, in particular, spatial (e.g. patterns involving place A or patterns where the movement direction is outwards from the centre) and temporal (e.g. patterns occurring in the morning).

Besides sequential patterns, node-link drawings superimposed on a map can also represent rules referring to specific places, for example:

- trafficJam (Pisa, 7:30AM) $\Longrightarrow$ trafficJam (Lucca, 8:30AM) or
- trafficJam (Pisa, t) $\Longrightarrow$ trafficJam (Lucca, t+1h)

(see Kuijpers et al., this volume). However, this approach will not work for rules involving more general spatial concepts such as "city centre" and "outskirts", "pedestrian area" and "major thoroughfare", etc., which have no precise localisation and/or crisp boundaries, and hence cannot be adequately represented on a map display. Instead, such concepts can be represented verbally or symbolically, for example, with the use of the system of signs, the so-called "choremes", suggested by Roger Brunet for the representation of spatial objects and relations (cited in [15]). It seems reasonable to study which form, symbolic or verbal, is more effective and convenient for users.

5.4 Visualisation of classification trees

Classification is applied to a set of records that contain class labels in order to create a profile for a member of each class from the values of available attributes. A typical result is a decision tree, which can be represented as a flow chart structure (as in Fig. 15) consisting of internal nodes, leaf nodes and branches. Each internal node represents a test on an attribute and each branch represents one of the results of that test. Each leaf node represents, ideally, a single class but in practice leaf nodes often represent several classes that could not be completely separated on the basis of available attribute values.

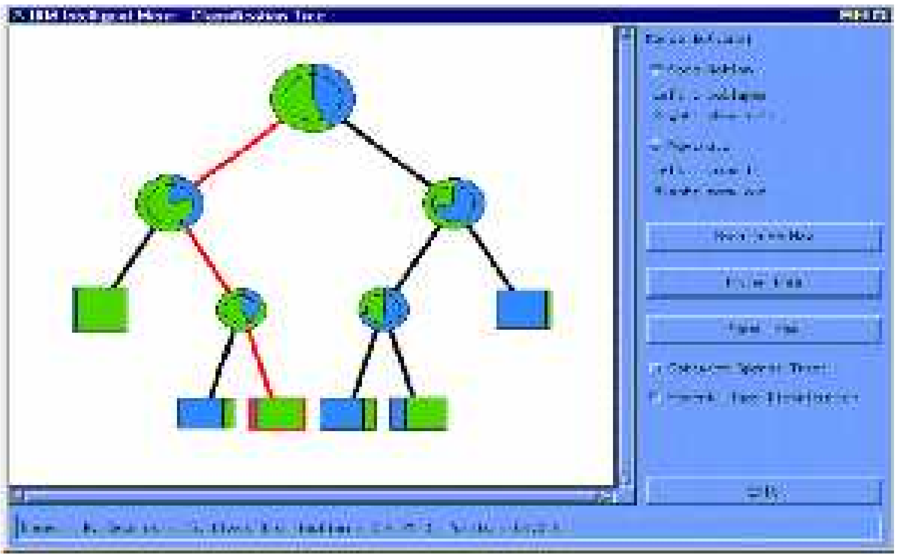

Fig. 15. A visualisation of a classification tree in the IBM Intelligent Miner software.

In movement data, class labels can be attached to records according to the types of the moving entities (e.g. pedestrians or vehicles) and their properties (e.g. people may be divided by age or occupation). The purpose of applying classification methods may be to find out how the classes of entities so defined differ by their movement characteristics. A user may also wish to divide trajectories or movement episodes into classes according to the movement direction, speed, length, or geographic position and to see how these classes differ in terms of other characteristics. In any case, the use of tree (i.e. node-link structure with specific properties) seems to be the most natural approach to the visualisation of the results. It is useful to extend this representation technique: where nodes refer to specific geographical locations or regions, interactive links to appropriate map displays should be included.

This is analogous to combining concept maps with multimedia displays [1] [10].

5.5 General notes

While specific research on visualisation of patterns that can be found in movement data is yet to be done, the following general considerations may provide some guidance:

1. Node-link structures are widely used and are therefore familiar to users and easily understood. Such structures are well suited to representation of patterns involving various kinds of relationships. In particular, they can represent sequence patterns in movement data as well as correlation and dependency patterns.
2. Node-link structures may incorporate various media, in particular, maps.
3. The use of maps is reasonable when patterns refer to specific geographical locations or regions.
4. Besides techniques for pattern visualisation, it is necessary to design tools for navigation through the set of patterns and for management of the patterns, which includes filtering, re-arrangement, and establishing of links.

6 Conclusion

Current state-of-the-art methods and tools for visual and interactive exploration of movement data have significant limitations regarding the volumes of data they can be applied to. In this chapter, we have outlined a road map to developing methods for visual analysis of massive datasets, with numerous moving entities and long time series of measurements. The methods are based on data aggregation, which is performed prior to the visualisation, as well as the use of computational analysis techniques. A number of technical problems need to be solved; in particular, effective linking between several displays presenting differently aggregated data.

The main goal of data exploration is detecting patterns and relationships in the data. We have considered the possible types of patterns an analyst may seek for in movement data. The role of interactive visual techniques is to allow the user to detect these patterns. We have also pointed out the need in tools for recording discovered patterns and in methods for the visualisation of patterns. Visualisation is necessary for a joint analysis of all detected patterns in order to gain an overall understanding of the data. This applies both to patterns detected by a human analyst and to patterns derived automatically, in particular, by means of data mining algorithms. Visualisation of patterns is also required when the analyst wishes to communicate his/her discoveries to others. Currently, the problem of pattern visualisation, in particular, visualisation of movement patterns, is far from being solved and requires further research efforts.

References

1. S. R. Alpert. Comprehensive mapping of knowledge and information resources: The case of webster. In *Knowledge and Information Visualization*, pages 220–237, 2005.
2. N. Andrienko and G. Andrienko. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer, 2006.
3. N. Andrienko, G.Andrienko, and P.Gatalsky. Impact of data and task characteristics on design of spatio-temporal data visualization tools. In *Exploring Geovisualization*, pages 201–222. Elsevier, 2005.
4. N. V. Andrienko, G. L. Andrienko, and P. Gatalsky. Supporting visual exploration of object movement. In *Advanced Visual Interfaces*, pages 217–220, 2000.
5. M. Q. W. Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Advanced Visual Interfaces*, pages 110–119, 2000.
6. J. Bertin. *Semiology of Graphics. Diagrams, Networks, Maps*. University of Wisconsin Press, 1983.
7. A. Buja, J. A. McDonald, J. Michalak, and W. Stuetzle. Interactive data visualization using focusing and linking. In *VIS '91: Proceedings of the 2nd conference on Visualization '91*, pages 156–163, Los Alamitos, CA, USA, 1991. IEEE Computer Society Press.
8. R. N. Buliung and P. S. Kanaroglou. An exploratory spatial data analysis (esda) toolkit for the analysis of activity/travel data. In A. Laganà, M. L. Gavrilova, V. Kumar, Y. Mun, C. J. K. Tan, and O. Gervasi, editors, *ICCSA (2)*, volume 3044 of *Lecture Notes in Computer Science*, pages 1016–1025. Springer, 2004.
9. P. Buono. Analysing association rules with an interactive graph-based technique. In *Proceedings of the HCI International Conference, Crete, Greece*, pages 675–679, 2003.
10. A. J. Cañas, R. Carff, G. Hill, M. M. Carvalho, M. Arguedas, T. C. Eskridge, J. Lott, and R. Carvajal. Concept maps: Integrating knowledge and information visualization. In *Knowledge and Information Visualization*, pages 205–219, 2005.
11. D. Carr. Looking at Large Data Sets Using Binned Data Plots. In *Computing and Graphics in Statistics*, pages 7–39. Springer, 1991.
12. I. Drecki and P. Forer. *Tourism in New Zealand - International Visitors on the Move (A1 Cartographic Plate)*. Tourism, Recreation Research and Education Centre (TRREC): Lincoln University, Lincoln, 2000.
13. J. A. Dykes and D. M. Mountain. Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications. *Computational Statistics & Data Analysis*, 43(4):581–603, 2003.
14. S. Eick. Engineering Perceptually Affective Visualizations for Abstract Data. In *Scientific Visualization Overviews, Methodologies and Techniques*, pages 191–210. IEEE Computer Science Press, 1997.
15. C. Elzakker. *The use of maps in the exploration of geographic data*. doctors dissertation, Utrecht : University of Utrecht (Netherlands Geographical Studies 326), 2004.
16. U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. *AI Magazine*, 17(3):37–54, 1996.

17. P. Forer and O. Huisman. Space, Time and Sequencing: Substitution at the Physical/Virtual Interface. In *Information, Place and Cyberspace: Issues in Accessibility*, pages 73–90. Springer, 2000.
18. M. Ganesh, E. Han, V. Kumar, S. Shekhar, and J. Srivastava. Visual data mining: Framework and algorithm development, 1996.
19. F. Giannotti, M. Nanni, and D. Pedreschi. Efficient mining of temporally annotated sequences. In *SDM*, 2006.
20. D. Guo, J. Chen, A. M. MacEachren, and K. Liao. A visualization system for space-time and multivariate patterns (vis-stamp). *IEEE Trans. Vis. Comput. Graph.*, 12(6):1461–1474, 2006.
21. D. Guo and M. Gahegan. Spatial ordering and encoding for geographic data mining and visualization. *J. Intell. Inf. Syst.*, 27(3):243–266, 2006.
22. M. Hao, M. Hsu, U. Dayal, S. Wei, T. Sprenger, and T. Holenstein. Market basket analysis visualization on a spherical surface. Technical Report http://www.hpl.hp.com/techreports/2001/HPL-2001-3.pdf, 2001.
23. B. Hetzler, W. Harris, S. Havre, and P. Whitney. Visualizing the full spectrum of document relationships, 1998.
24. S. Imfeld. *Time, points and space: Analysis of wildlife data in GIS*. Dissertation, University of Zurich, Department of Geography, Zurich, 2000.
25. T. Kapler and W. Wright. Geotime information visualization. *Information Visualization*, 4(2):136–146, 2005.
26. D. A. Keim. Information visualization and visual data mining. *IEEE Trans. Vis. Comput. Graph.*, 8(1):1–8, 2002.
27. I. Kopanakis. *Visualization of Data Mining Outcomes*. http://www.csd.uoc.gr/k̃opanak/Sources/pattern_vis_review.pdf, 2006.
28. I. Kopanakis and B. Theodoulidis. Visual data mining modeling techniques for the visualization of mining outcomes. *J. Vis. Lang. Comput.*, 14(6):543–589, 2003.
29. M.-J. Kraak. The space-time cube revisited from a geovisualization perspective. In *21st International Cartographic Conference, Durban, South-Africa*, pages 1988–1995, 2003.
30. Y. Kurata and M. J. Egenhofer. Structure and semantics of arrow diagrams. In *COSIT*, pages 232–250, 2005.
31. M.-P. Kwan and J. Lee. Geovisualization of human activity patterns using 3-D GIS: a time-geographic approach. In *Spatially Integrated Social Science*. Oxford University Press, 2004.
32. P. Laube, S. Imfeld, and R. Weibel. Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science*, 19(6):639–668, 2005.
33. H. Miller. Modeling Accessibility Using Space-Time Prism Concepts within Geographical Information Systems: Fourteen Years . In *Classics of IJGIS*, pages 177–182. CRC Press, 2006.
34. D. Mountain. *Exploring mobile trajectories: An investigation of individual spatial behaviour and geographic filters for information retrieval.* Dissertation, City University, London, 2005.
35. D. Mountain. Visualizing, querying and summarizing individual spatio-temporal behaviour. In *Exploring Geovisualization*, pages 181–200. Elsevier, 2005.
36. D. Mountain and J. Dykes. What i did on my vacation: Spatio-temporal log analysis with interactive graphics and morphometric surface derivatives. In *Proceedings of GISRUK*, 2002.

37. D. Mountain and J. Raper. Modelling human spatio-temporal behaviour: A challenge for location-based services. In *Geocomputation'2001 - 6th International Conference on Geocomputation*, 2001.
38. C. Newton. Graphics: from alpha to omega in data analysis. In *Graphical Representation of Multivariate Data*, pages 59–92. Academic Press, 1978.
39. C. North and B. Schneiderman. A taxonomy of multiple window coordinations. Technical Report CS-TR-3854, 1997.
40. J. C. Roberts. On encouraging multiple views for visualisation. In *Information Visualization*, Los Alamitos, CA, USA, 1998. IEEE Computer Society.
41. B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *IEEE Visual Languages*, number UMCP-CSD CS-TR-3665, pages 336–343, College Park, Maryland 20742, U.S.A., 1996.
42. T. Slocum, R. MacMaster, F. Kessler, and H. Howard. *Thematic Cartography and Geographic Visualization*. Prentice Hall, 2005.
43. R. Spence and L. Tweedie. The attribute explorer: information synthesis via exploration. *Interacting with Computers*, 11(2):137–146, 1998.
44. S.-O. Tergan and T. Keller. *Knowledge and Information Visualization: Searching for Synergies*. Springer, 2005.
45. J. Thomas and K. Cook. *Illuminating the Path. The Research and development Agenda for Visual Analytics*. IEEE Computer Society, 1983.
46. W. Tobler. Experiments in migration mapping by computer. *The American Cartographer*, 14(2):155–163, 1987.
47. W. Tobler. *Display and Analysis of Migration Tables*. http://www.geog.ucsb.edu/t̃obler/presentations/shows/A_Flow_talk.htm, 2005.
48. B. Tversky, J. B. Morrison, and M. Bétrancourt. Animation: can it facilitate? *Int. J. Hum.-Comput. Stud.*, 57(4):247–262, 2002.
49. L. Wilkinson. *The grammar of graphics*. Springer-Verlag New York, Inc., New York, NY, USA, 1999.
50. H. Yu. Spatial-temporal gis design for exploring interactions of human activities. *Cartography and Geographic Information Science*, 33(1):3–19, 2006.
51. K. Zhao and B. Liu. Visual analysis of the behavior of discovered rules. In *ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD 2001), Proc. Workshop on Visual Data Mining*, pages 59–64, 2001.