Movement Patterns in Spatio-Temporal Data

Joachim Gudmundsson, National ICT Australia Ltd, Australia
Patrick Laube, University of Melbourne, Australia
Thomas Wolle, National ICT Australia Ltd, Australia

SYNONYMS
Motion patterns; Trajectory patterns

DEFINITION

Spatio-temporal data is any information relating space and time. This entry specifically considers data involving point objects moving over time. The terms *entity* and *trajectory* will refer to such a point object and the representation of its movement, respectively. *Movement patterns* in such data refer to (salient) events and episodes expressed by a set of entities.

In the case of moving animals, movement patterns can be viewed as the spatio-temporal expression of behaviours, as for example in flocking sheep or birds assembling for the seasonal migration. In a transportation context, a movement pattern could be a traffic jam.

Only formalised patterns are detectable by algorithms. Hence, movement patterns are modelled as any arrangement of subtrajectories that can be sufficiently defined and formalised, see for example the patterns illustrated in Figure 1. A pattern usually involves a certain number of entities. Furthermore a pattern starts and ends at certain times (temporal footprint), and it might be restricted to a subset of space (spatial footprint).

HISTORICAL BACKGROUND

The analysis of movement patterns in spatio-temporal data is for two main reasons a relatively young and little developed research field. First, emerging from static cartography, geographical information science and theory struggled for a long time with the admittedly substantial challenges of handling dynamics. For many years, occasional changes in a cadastral map were challenging enough, not to mention the constant change of location as is needed for modelling movement.

Second, only in recent years has the technological advancement in tracking technology reached a level that allowed the seamless tracking of individuals needed for the analysis of movement patterns. For many years, the tracking of movement entities has been a very cumbersome and costly undertaking. Hence, movement patterns could only be addressed for single individuals or very small groups. Hägerstrand's time geography [10] may serve as a starting point of a whole branch of geographical information science representing individual trajectories in 3D. The two spatial dimensions combined with an orthogonal temporal axis proved to be a very
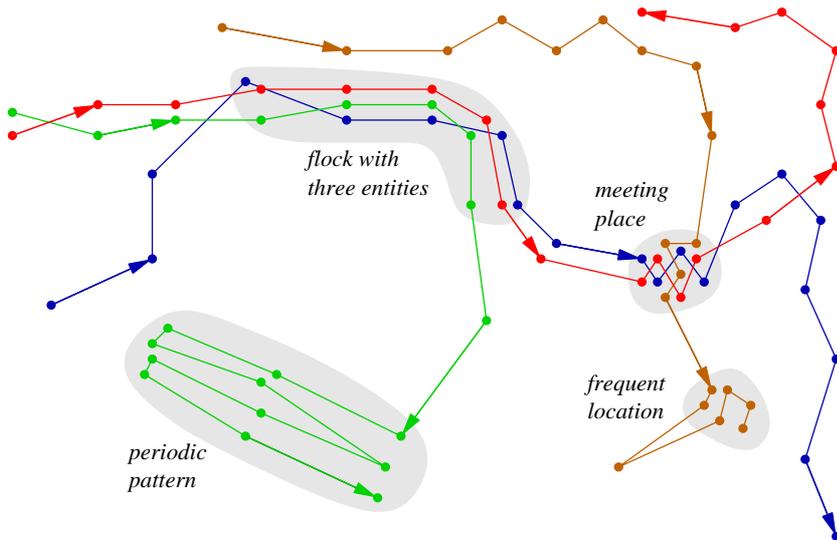
Figure 1: Illustrating the trajectories of four entities moving over 20 time steps. The following patterns are highlighted: a *flock* of three entities over five time-steps, a *periodic pattern* where an entity shows the same spatio-temporal pattern with some periodicity, a *meeting place* where three entities meet for four time steps, and finally, a *frequently visited location* which is a region where a single entity spends a lot of time.

powerful concept for exploring various kinds of spatio-temporal relationships, including movement patterns.

With GPS and various other tracking technologies movement pattern research entered a new era, stepping from 'thread trailing' and 'mark and recapture' approaches to low cost, almost continuous capture of individual trajectories with possibly sub-second sampling rates. Within a few years the situation completely reversed from a notorious data deficit to almost a data overkill, with a lack of suited analytical concepts coping with the sudden surge of movement data. Consequently, the huge potential of analysing movement in spatio-temporal data has recently attracted the interest of many research fields, both in theory and application, as is outlined in the next two sections.

SCIENTIFIC FUNDAMENTALS

Assume that the entities in Figure 1 are sheep on a pasture and that they are observed by a geographer, a database expert and a computational geometer. Even though all three experts see the very same sheep, they may all perceive totally different things. The geographer might interpolate a sheep density surface of the pasture. For the database expert in contrast, each sheep may represent a leaf in a dynamic tree optimised for fast queries. Finally, the computational geometer might triangulate the sheep locations in order to detect a flocking pattern. Even though the sheep will not care, their grazing challenges various research fields handling spatio-temporal data. The following overview bundles the different perspectives addressing movement patterns into the three sections exploration, indexing and data mining. See Figure 2 for a comprehensible access guide to recommended reading.

GIScience: exploratory data analysis and visualisation In GIScience the term 'pattern' is used in various contexts and meanings when addressing movement. However, as a common denominator, movement patterns are generally conceptualised as salient movement events or

episodes in the geospatial representation of moving entities. Given GIScience's legacy in cartography, it is not surprising that movement patterns are often addressed by a combination of geovisualisation and exploration. Exploratory analysis approaches combine the speed and patience of computers with the excellent capability of humans to detect the expected and discover the unexpected given an appropriate graphical representation.

Salient movement patterns may emerge from (i) two-dimensional maps of fixes aligned in trajectories, (ii) movie-like animated maps or even (iii) three-dimensional representations of movement, if time is used as a third, orthogonal axis.

(i) Basic movement patterns are obvious from simple plotting of movement trajectories on a two-dimensional map. Trajectories bundled in narrow, directed bottlenecks represent often used corridors. Less focussed trajectory footprints represent more arbitrary movement, such as in grazing animals or visitors at a sports event strolling around a stadium. The application of GIS analysis tools on points and lines representing moving entities has proven to be a very effective approach. For example, GIS tools for generalisation, interpolation and surface generation may be applied to support the detection of movement patterns in trajectory data. Brillinger et al. [5] use a regularly sampled vector field to illustrate the overall picture of animals moving in their habitat, with each vector coding in orientation and size for mean azimuth and mean speed at that very location. Dykes and Mountain [8] use a continuous density surface and a 'spotlight' metaphor for the detection of activity patterns. Again, common GIS tools such as algorithms initially designed for the analysis of digital terrain models can easily be adopted for the search for salient movement patterns, for instance to identify 'peaks' of frequent visitation and 'ridges' of busy corridors [8].

(ii) Animation is suited to uncover specific movement behaviours of individuals and groups. Animating moving entities with a constant moving time window in the so-called dynamic view uncovers speed patterns of individuals [2, 8]. Flocking or converging are more complex patterns of coordination in groups. Such group patterns are very striking when animating even large numbers or individuals in a movie-like animation.

(iii) The extension of a two-dimensional map with a third orthogonal time axis produces a very powerful tool for uncovering movement patterns. Such ideas go back to Hägerstrand's time geography [10] and have often been adopted in present day geocomputation [12]. In the specific geometry in such a three-dimensional space-time aquarium episodes of immobility and certain speed behaviours produce distinctive patterns of vertical and inclined time lines, respectively. Furthermore, patterns of spatio-temporal collocation can be identified from vertical bottleneck structures in sets of time lines [12].

Indexing spatio-temporal trajectories In the database community considerable research has been focussing on spatial and temporal databases. Research in the spatio-temporal area in many ways started with the dissertations by Lorentzos [15] in 1988 and Langran [13] in 1989. Not surprisingly research has mainly focussed on indexing databases so that basic queries concerning the data can be answered efficiently. The most common queries considered in the literature are variants of nearest neighbour queries and range searching queries. For example:

- Spatio-temporal range query, e.g. 'Report all entities that visited region S during the time interval $[t_1, t_2]$.'

3

- Spatial nearest neighbours given a time interval, e.g. 'Report the entity closest to point p at time t.'

- Temporal nearest neighbours given a spatial region, e.g. 'Report the first entity visiting region S.'

In general one can classify indexing methods used for spatio-temporal data into Parametric Space Indexing methods (PSI) and Native Space Indexing methods (NSI). The PSI method uses the parametric space defined by the movement parameters, and is an efficient approach especially for predictive queries. A typical approach, described by Săltenis et al. [17] is to represent movement defined by its velocity and projected location along each spatial dimension at a global time reference. The parametric space is then indexed by a new index structure referred to as the TPR-tree (Time Parameterised R-tree). The TPR-tree is a balanced, multi-way tree with the structure of an R-tree. Entries in leaf nodes are pairs of the position of a moving point and a pointer to the moving point, and entries in internal nodes are pairs of a pointer to a subtree and a rectangle that bounds the positions of all moving points or other bounding rectangles in that subtree. The position of a moving point is represented by a reference position and a corresponding velocity vector. To bound a group of d-dimensional moving points, d-dimensional bounding rectangles are used that are also time parameterised, i.e. their coordinates are functions of time. A time-parameterised bounding rectangle bounds all enclosed points or rectangles at all times not earlier than the current time. The search algorithm for a range query also performs computation on the native space by checking the overlap between the range of the query and the trapezoid representation of the node.

The NSI methods represent movement in d dimensions as a sequence of line segments in $d+1$ dimensions, using time as an additional dimension, see for example the work by Hadjieleftheriou et al. [9]. A common approach is to use a multi-dimensional spatial access method like the R-tree. An R-tree would approximate the whole spatio-temporal evolution of an entity with one Minimum Bounding Region (MBR) that tightly encloses all the locations occupied by the entity during its lifetime. An improvement for indexing movement trajectories is to use a multi version index, like the Multi Version R-tree (MVR-tree), also known as a persistent R-tree. This index stores all the past states of the data evolution and allows updates to the most recent state. The MVR-tree divides long-lived entities into smaller intervals by introducing a number of entity copies. A query is directed to the exact state acquired by the structure at the time that the query refers to; hence, the cost of answering the query is proportional to the number of entities that the structure contained at that time.

Algorithms and data mining In the previous section different indexing approaches were discussed. This section will focus on mining trajectories for spatio-temporal patterns. This has mainly been done using algorithmic or data mining approaches.

The most popular tools used in the data mining community for spatio-temporal problems has been association rule mining (ARM) and various types of clustering. Association rule mining seeks to discover associations among transactions within relational databases. An association rule is of the form $X \implies Y$ where X (antecedents) and Y (consequents) are disjoint conjunctions of attribute-value pairs. ARM uses the concept of *confidence* and *support*. The confidence of the rule is the conditional probability of Y given X, and the support of the rule is the prior

probability of X and Y.

The probability is usually the observed frequency in the data set. Now the ARM problem can be stated as follows. Given a database of transactions, a minimal confidence threshold and a minimal support threshold, find all association rules whose confidence and support are above the corresponding thresholds.

Chawla and Verhein [18] defined spatio-temporal association rules (STARs) that describe how entities move between regions over time. They assume that space is partitioned into regions, which may be of any size and shape. The aim is to find interesting regions and rules that predict how entities will move through the regions. A region is interesting when a large number of entities leaves (sink), a large number of entities enters (source) or a large number of entities enters and leaves (thoroughfare).

A STAR $(r_i, T_1, q) \Longrightarrow (r_j, T_2)$ denotes a rule where entities in a region r_i satisfying condition q during time interval T_1 will appear in region r_j during time interval T_2. The support of a rule δ is the number, or ratio, of entities that follow the rule. The *spatial* support takes the size of the involved regions into consideration. That is, a rule with support s involving a small region will have a larger spatial support than a rule with support s involving a larger region. Finally, the confidence of a rule δ is the conditional probability that the consequent is true given that the antecedent is true. By traversing all the trajectories all possible movements between regions can be modelled as a rule, with a spatial support and confidence. The rules are then combined into longer time intervals and more complicated movement patterns.

Some of the most interesting spatio-temporal patterns are periodic patterns, e.g. yearly migration patterns or daily commuting patterns. Mamoulis et al. [16] considered the special case when the period is given in advance. They partition space into a set of regions which allows them to define a pattern P as a τ-length sequence of the form $r_0, r_1, \ldots, r_{\tau-1}$, where r_i is a spatial region or the special character *, indicating the whole spatial universe. If the entity follows the pattern enough times, the pattern is said to be *frequent*. However, this definition imposes no control over the density of the regions, i.e. if the regions are too large then the pattern may always be frequent. Therefore an additional constraint is added, namely that the points of each subtrajectory should form a cluster inside the spatial region.

Kalnis et al. [11] define and compute moving clusters where entities might leave and join during the existence of a moving cluster. For each fixed discrete time-step t_i they use standard clustering algorithms to find clusters with a minimum number of entities and a minimum density. Then they compare any cluster c found for t_i with any (moving) cluster c' found for time-step t_{i-1}. If c and c' have enough entities in common, which is formally specified by a threshold value, then c' can be extended by c, which results in a *moving cluster*. They propose several ideas to increase the speed of their method, e.g. by avoiding redundant cluster comparisons, or approximating moving clusters instead of giving exact solutions, and they experimentally analyse their performance.

In 2004 Laube et al. [14] defined a collection of spatio-temporal patterns based on direction of movement and location, e.g. flock, leadership, convergence and encounter, and they gave algorithms to compute them efficiently. As a result there were several subsequent articles studying the discovery of these patterns. Benkert et al. [4] modified the original definition of a flock to be a set of entities moving close together during a time interval. Note that in this definition the

entities involved in the flock must be the same during the whole time interval, in contrast to the moving cluster definition by Kalnis et al. [11]. Benkert et al. [4] observed that a flock of m entities moving together during k time steps corresponds to a cluster of size m in $2k$ dimensional space. Thus the problem can be restated as clustering in high dimensional space. To handle high dimensional space one can use well-known dimensionality reduction techniques. There are several decision versions of the problem that have been shown to be NP-hard, for example deciding if there exists a flock of a certain size, or of a certain duration. The special case when the flock is stationary is often called a *meeting* pattern.

Andersson et al. [1] gave a more generic definition of the pattern *leadership* and discussed how such leadership patterns can be computed from a group of moving entities. The proposed definition is based on behavioural patterns discussed in the behavioural ecology literature. The idea is to define a leader as an entity that (1) does not follow anyone else, (2) is followed by a set of entities and (3) this behaviour should continue for a duration of time. Given these rules all leadership patterns can be efficiently computed.

Be it exploratory analysis approaches, indexing techniques or data mining algorithms, all effort put in theory ultimately leads to more advanced ways of inferring high level process knowledge from low level tracking data. The following section will illustrate a wide range of fields where such fundamentals underlie various powerful applications.

KEY APPLICATIONS

Animal Behaviour The observation of behavioural patterns is crucial to animal behaviour science. So far, individual and group patterns are rather directly observed than derived from tracking data. However, there are more and more projects that collect animal movement by equipping them with GPS-GSM collars. For instance, since 2003 the positions of 25 elks in Sweden are obtained every 30 minutes. Other researchers attached small GPS loggers to racing pigeons and tracked their positions every second during a pigeon's journey. It is even possible to track the positions of insects, e.g. butterflies or bees, however most of the times non-GPS based technologies are used that allow for very small and light sensors or transponders. Analysing movement patterns of animals can help to understand their behaviour in many different aspects. Scientists can learn about places that are popular for individual animals, or spots that are frequented by many animals. It is possible to investigate social interactions, ultimately revealing the social structure within a group of animals. A major focus lies on the investigation of leading and following behaviour in socially interacting animals, such as in a flock of sheep or a pack of wolves [7]. On a larger scale, animal movement data reflects very well the seasonal or permanent migration behaviour. In the animation industry, software agents implement movement patterns in order to realistically mimic the behaviour of animal groups. Most prominent is the flocking model implemented in NetLogo which mimics the flocking of birds [19].

Human Movement Movement data of people can be collected and used in several ways. For instance, using mobile phones that communicate with a base station is one way to gather data about the approximate locations of people. Traffic-monitoring devices such as cameras can deliver data on the movement of vehicles. With the technological advancement of mobile and position

aware devices, one could expect that tracking data will be increasingly collectable. Although tracking data of people might be available in principle, ethical and privacy aspects need to be taken into consideration before gathering and using this data [6]. Nonetheless, if the data is available, it could be used for urban planning, e.g. to plan where to build new roads or where to extend public transport.

The detection of movement patterns can furthermore be used to optimise the design of location-based-services (LBS). The services offered to a moving user could not only be dependent on the actual position, but also on the estimated current activity, which may be derived from a detected movement pattern.

Traffic Management Movement patterns are used for traffic management in order to detect undesirable or even dangerous constellations of moving entities, such as traffic jams or aeroplane course conflicts. Traffic management applications may require basic Moving Object Database queries, but also more sophisticated movement patterns involving not just location but also speed, movement direction and other activity parameters.

Surveillance and Security Surveillance and intelligence services might have access to more detailed data sets capturing the movement of people, e.g. coordinates from mobile phones or credit card usage, video surveillance camera footage or maybe even GPS data. Apart from analysing the movement data of a suspect to help prevent further crime, it is an important task to analyse the entire data set to identify suspicious behaviour in the first place. This leads to define 'normal behaviour' and then search the data for any outliers, i.e. entities that do not show normal behaviour. Some specific activities and the corresponding movement patterns of the involved moving entities express predefined signatures that can be automatically detected in spatio-temporal or footage data. One example is that fishing boats in the sea around Australia have to report their location in fixed intervals. This is important for the coast guards in case of an emergency, but the data can also be used to identify illegal fishing in certain areas. Another example is that a car thief is expected to move in a very characteristic and hence detectable way across a surveilled car park. Movement patterns have furthermore attracted huge interests in the field of spatial intelligence and disaster management. Batty et al. [3] investigated local pedestrian movement in the context of disaster evacuation where movement patterns such as congestion or crowding are key safety issues.

Military and Battlefield The digital battlefield is an important application of moving object databases. Whereas real-time location data of friendly troops is easily accessible, the enemy's location may be obtained from reconnaissance planes with only little time lag. Moving object databases not only allow the dynamic updating of location and status of tanks, aeroplanes and soldiers, but also answering spatio-temporal queries and detecting complex movement patterns. Digital battlefield applications answer spatio-temporal range queries like 'Report all friendly tanks that are currently in region S.' A more complex movement pattern in a digital battlefield context would be the identification of the convergence area where the enemy is currently concentrating his troops.

Sports Scene Analysis Advancements in many different areas in technology are also influencing professional sports. For example, some of the major tennis tournaments provide three-dimensional reconstructions of every single point played, tracking the players and the balls. It is furthermore known that, e.g. football coaches routinely analyse match video archives to learn about an opponents behaviours and strategies. Making use off tracking technology, the movement of the players and the ball can be described by 23 trajectories over the length of the match. Researchers were able to develop a model that is based on the interactions between the players and the ball. This model can be used to quantitatively express the performance of players, and more general, it might lead to an improved overall strategy. Finally, real-time tracking systems are developed that keep track of both players and the ball in order to assist the referee with the detection of the well-defined but nevertheless hard to perceive offside pattern.

Movement in Abstract Spaces In contrast to tracking and analysing the movement of animals and people on the surface of the earth, it is also possible to obtain and analyse spatio-temporal data in abstract spaces also in higher dimensions. Every scatter plot that constantly updates the changes in the x and y values, produces individual trajectories open for movement analysis. Two stock exchange series plotted against each other could build such a dynamic scatter-plot. As another example, basic ideological conflicts can be used to construct abstract ideological spaces. Performing factor analysis on referendum data, researchers hypothesised a structure of mentality consisting of dimensions such as 'political left vs. political right' or 'liberal vs. conservative'. Whole districts or even individuals such as members of parliament could now be localised and re-localised in such ideological space depending on their voting behaviour and its change over time, respectively. Movement in such a space represents the change of opinions and analysing this can lead to more insight and understanding of human psychology and politics.

FUTURE DIRECTIONS

For simplicity reasons, theory and application of movement patterns in spatio-temporal data focussed so far largely on moving point objects. However, many processes can only be modelled as dynamics in fields or in their discretised counterparts that is dynamic polygons. When monitoring a hurricane threatening urban areas, the tracking of its eye alone may not provide sufficient information, but additional tracking of its changing perimeter will be required. The consideration of both location and change of polygonal objects raises the conceptualisation and detection of movement patterns to a higher level of complexity, which has only rarely been addressed so far.

For the many fields interested in movement, the overall challenge lies in relating movement patterns with the underlying geography, in order to understand where, when and ultimately why the entities move the way they do. Grazing sheep, for example, may perform a certain movement pattern only when they are on a certain vegetation type. Homing pigeons may show certain flight patterns only when close to a salient landscape feature such as a river or a highway. And, the movement patterns expressed by a tracked vehicle will obviously be very dependant on the environment the vehicle is moving in, be it in a car park, in a suburb or on a highway. Thus, patterns have to be conceptualised that allow linking of the movement with the embedding environment.
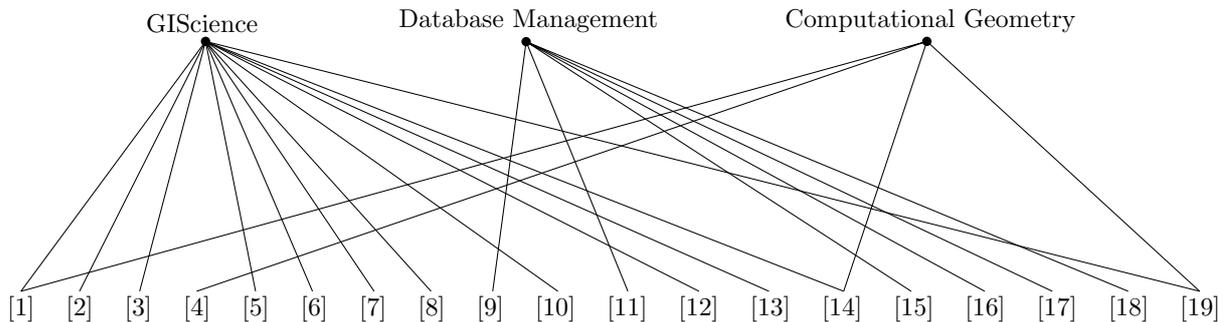
Figure 2: Access guide to the references in the further reading section below.

Cross References

Computational Geometry (Ch. 7), Terrain Modeling (Ch. 15), Query Operation & Query Languages (Ch. 30), Navigation and Transportation Issues (Ch. 43), Management of Moving Objects (Ch. 45), Privacy in location-based services (Ch. 47), Location Based Services (Ch. 73), GIS for Transportation (Ch. 74), GIS in Army (Ch. 76), GIS Intelligence (Ch. 79), Legal and Ethical Issues of GIS. Association Rule Mining (Ch. 80), Spatio-temporal modeling and reasoning (Ch. 83), Spatio-temporal Indexing (Ch. 85), Spatio-Temporal Representation (Ch. 88).

RECOMMENDED READING

[1] M. Andersson, J. Gudmundsson, P. Laube, and T. Wolle. Reporting leadership patterns among trajectories. In *Proceedings of the 22nd ACM Symposium on Applied Computing*, 2007.

[2] N. V. Andrienko and G. L. Andrienko. Interactive maps for visual data exploration. *International Journal of Geographical Information Science*, 13(4):355–374, 2003.

[3] M. Batty, J. Desyllas, and E. Duxbury. The discrete dynamics of small-scale spatial events: agent-based models of mobility in carnivals and street parades. *International Journal of Geographical Information Science*, 17(7):673–697, 2003.

[4] M. Benkert, J.Gudmundsson, F. Hübner, and T. Wolle. Reporting flock patterns. In *Proceedings of the 14th European Symposium on Algorithms*, volume 4168 of *Lecture Notes in Computer Science*, pages 660–671. Springer, 2006.

[5] D. R. Brillinger, H. K. Preisler, A. A. Ager, and J. G. Kie. An exploratory data analysis (EDA) of the paths of moving animals. *Journal of statistical planning and inference*, 122(2):43–63, 2004.

[6] J. E. Dobson and P. F. Fisher. Geoslavery. *IEEE Technology and Society Magazine*, 22(1):47–52, 2003.

[7] B. Dumont, A. Boissy, C. Achard, A. M. Sibbald, and H. W. Erhard. Consistency of animal order in spontaneous group movements allows the measurement of leadership in a group of grazing heifers. *Applied Animal Behaviour Science*, 95(1–2):55–66, 2005.

[8] J. A. Dykes and D. M. Mountain. Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and application. *Computational Statistics and Data Analysis*, 43(4):581–603, 2003.

[9] M. Hadjieleftheriou, G. Kollios, V. J. Tsotras, and D. Gunopulos. Indexing spatio-temporal archives. *The VLDB Journal*, 15(2):143–164, 2006.

[10] T. Hägerstrand. What about people in regional science. *Papers of the Regional Science Association*, 24:7–21, 1970.

[11] P. Kalnis, N. Mamoulis, and S. Bakiras. On discovering moving clusters in spatio-temporal data. In C. B. Medeiros, M. J. Egenhofer, and E. Bertino, editors, *Proceedings of the 9th International Symposium on Advances in Spatial and Temporal Databases*, volume 3633 of *Lecture Notes in Computer Science*, pages 364–381. Springer, 2005.

[12] M. P. Kwan. Interactive geovisualization of activity-travel patterns using three dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research Part C*, 8(1–6):185–203, 2000.

[13] G. Langran. *Time in Geographic Information Systems*. PhD thesis, University of Washington, 1999.

[14] P. Laube, M. van Kreveld, and S. Imfeld. Finding REMO - detecting relative motion patterns in geospatial lifelines. In P. F. Fisher, editor, *Developments in Spatial Data Handling*, Proceedings of the 11th International Symposium on Spatial Data Handling, pages 201–214. Springer, Berlin Heidelberg, DE, 2004.

[15] N. A. Lorentzos. *A Formal Extension of the Relational Model for the Representation and Manipulation of Generic Intervals*. PhD thesis, Birbeck College, University of London, 1988.

[16] N. Mamoulis, H. Cao, G. Kollios, M. Hadjieleftheriou, Y. Tao, and D. Cheung. Mining, indexing, and querying historical spatiotemporal data. In *Proceedings of the 10th ACM International Conference On Knowledge Discovery and Data Mining*, pages 236–245. ACM, 2004.

[17] S. Săltenis, C. S. Jensen, S. T. Leutenegger, and M. A. Lopez. Indexing the positions of continuously moving objects. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 331–342, 2000.

[18] F. Verhein and S. Chawla. Mining spatio-temporal association rules, sources, sinks, stationary regions and thoroughfares in object mobility databases. In *Proceedings of the 11th International Conference on Database Systems for Advanced Applications*, volume 3882 of *Lecture Notes in Computer Science*, pages 187–201. Springer, 2006.

[19] U. Wilensky. Netlogo flocking model. http://ccl.northwestern.edu/netlogo/models/Flocking.