Universität
Augsburg
University

# A Framework for Analysing Trajectories of Movement in a Dynamic Geographic Context

**Jean Damascène Mazimpaka**

Dissertation
for the degree of
Doctor of Natural Sciences (Dr. rer. nat.)
in the
Faculty of Applied Computer Science
University of Augsburg

May 2017

| | |
|---|---|
| Evaluation committee: | Prof. Dr. Sabine Timpf |
| | Dr. Urška Demšar |
| | Prof. Dr. Jukka M. Krisp |
| Date of defence: | May 26th, 2017 |

To my parents

# Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Dr. Sabine Timpf. By accepting me as a PhD student you enabled me to come to Germany and start the journey towards my longstanding dream. Without your tremendous guidance throughout, this journey could not have led to realising the dream. I would also like to thank Prof. Dr. Jukka M. Krisp for very constructive comments and recommendations during the period of my PhD studies.

I am so grateful to the German Academic Exchange Service (DAAD) for funding my PhD studies. I would like to thank Nicole Karrais for her support in my early days in Augsburg. My thanks also go to Dr. Christa Rottscheidt-Kleine for her support in Augsburg. I am thankful to Andreas Keler we shared office. Our discussion of issues common to our research topics did not only lead to co-authorship of papers but was also helpful on every day of my research. I am also thankful to all colleagues in our Geoinformatics Group: Eva Nuhn, Dr. Linfang Ding, Johannes Schwer, Dr. David Jonietz, and Dr. Carolin von Groote-Bidlingmaier for their helpful tips. Your company and support made my environment enjoyable.

I would like to thank other researchers we exchanged during my studies. Especially, I thank Dr. Urška Demšar and Prof. Dr. Nico Van de Weghe with whom I had discussions which contributed a lot in shaping my research topic.

I thank God for these wonderful people I met in my life and the good experience in Augsburg. Last but not least, I would like to thank my family for their love and encouragement.

# Abstract

Recent advances in location-aware technologies enable the collection of trajectories of moving entities, which can be useful in different application domains such as urban planning, transportation and environment management. The analysis of these trajectories has mainly focused on discovering movement patterns. However, the usefulness of the discovered patterns depends on the possibility to interpret and understand them. Recent studies have shown that the consideration of the movement context, while analysing trajectories, has the potential to support the understanding of movement patterns. However, the integration of movement context into the analysis of trajectories is still in its infancy and most of the available work considers only a static geographic context. This thesis develops a comprehensive conceptual and methodological framework for integrating a dynamic geographic context into the analysis of trajectories.

In the first step, a conceptual model relating the movement to its dynamic geographic context is developed. The thesis establishes a classification of geographic context elements and then proposes a set of qualitative relations, termed movement interactions, between the movement and the context. In the second step, the thesis proposes an analysis framework which exploits the conceptual model developed. The analysis framework is based on the process of Knowledge Discovery in Database (KDD). The thesis focuses on two steps, which correspond to the steps of the KDD process aimed at discovering and interpreting patterns. The first step applies data mining and spatial analysis methods to extract interactions from trajectories and context data. The second step quantifies the extracted interactions and explores the correlation or dependence between the quantified interactions and dynamic attributes of the movement and the context.

In order to evaluate the framework developed, the thesis executes three experiments using real trajectories of vehicle movement in urban environment. Each experiment focusses on specific challenges addressed by the thesis. The first experiment focuses on the temporal dynamics of the dynamic geographic context while the second experiment focusses on its spatial dynamics. While the first two experiments involve context data in pattern discovery, the third experiment involves context data for post-processing already discovered patterns. The experiments show that the integration of context data supports not only the interpretation of movement patterns but also a deeper understanding of the movement context. Furthermore, the experiments show that context data can be integrated at the pattern discovery stage or for post-processing already discovered patterns. The choice of the integration step depends on the data being analysed and the type of patterns being mined.

# Zusammenfassung

Aktuelle Errungenschaften in standortbezogenen Technologien ermöglichen Trajektorien Erfassung von sich beweglichen Objekten. Bewegungstrajektorien können in verschiedenen Anwendungsfeldern wie Stadtplanung, Transport und Umweltmanagement nützlich sein. Die Analyse dieser Trajektorien konzentriert sich vor allem auf die Detektion von Bewegungsmustern. Der Nutzen der detektierten Muster hängt von Möglichkeiten ab, diese zu interpretieren und zu verstehen. Neueste Studien zeigen, dass die Betrachtung des Bewegungskontextes bei der Analyse von Trajektorien das Verständnis von Bewegungsmustern unterstützt. Allerdings ist die Integration des Bewegungskontextes in die Analyse der Trajektorien noch nicht viel erforscht und die meisten verfügbaren Studien betrachten nur einen statischen geographischen Kontext. Diese Arbeit entwickelt ein umfassendes konzeptionelles und methodologisches Framework für die Integration eines dynamischen geografischen Kontexts in die Analyse von Trajektorien.

Im ersten Teil wird ein konzeptionelles Modell entwickelt, das Bewegung mit ihrem zugrundeliegenden dynamischen geographischen Kontext verknüpft. Die vorliegende Dissertation stellt eine Klassifikation der geographischen Kontextelemente vor und schlägt anschließend eine Reihe von qualitativen Beziehungen zwischen der Bewegung und dem Kontext vor. Diese qualitativen Beziehungen werden als Bewegungsinteraktionen bezeichnet. Im zweiten Teil der Arbeit wird ein Analyse-Framework vorgeschlagen, das auf dem entwickelten konzeptionellen Modell beruht. Das Analyse-Framework basiert auf dem Prozess der Knowledge Discovery in Database (KDD). Die Arbeit konzentriert sich auf zwei Schritte des KDD-Prozesses, die darauf abzielen, Muster zu detektieren und zu interpretieren. Der erste Schritt besteht aus Data Mining und räumlichen Analysemethoden mit dem Ziel Interaktionen aus Trajektorien und Kontextdaten zu extrahieren. Der zweite Schritt quantifiziert die extrahierten Interaktionen und untersucht die Korrelation oder Abhängigkeit zwischen den quantifizierten Interaktionen und den dynamischen Attributen der Bewegung und des Kontextes.

Um das entwickelte Framework zu bewerten, werden in der Arbeit drei Experimente durchgeführt, die erfasste Trajektorien von Fahrzeugbewegungen in urbanen Gebi-

eten benutzen. Jedes Experiment konzentriert sich auf spezifische Herausforderungen, mit denen sich diese Dissertation beschäftigt. Das erste Experiment konzentriert sich auf die zeitliche Dynamik des dynamischen geographischen Kontextes, während das zweite Experiment auf seine räumliche Dynamik fokussiert. Während die ersten beiden Experimente Kontextdaten vor der Extrahierung von Bewegungsmustern integrieren, integriert das dritte Experiment Kontextdaten für die Nachbearbeitung bereits extrahierter Bewegungsmuster. Die Experimente zeigen, dass die Integration von Kontextdaten nicht nur die Interpretation von Bewegungsmustern unterstützt, sondern auch ein tieferes Verständnis des Bewegungskontextes ermöglicht. Darüber hinaus zeigen die Experimente, dass Kontextdaten entweder bei der Extrahierung von Bewegungsmustern oder bei der Nachbearbeitung bereits extrahierter Bewegungsmuster integriert werden können.

# Table of contents

# List of figures

# List of tables

# Chapter 1

# Introduction

## 1.1 Motivation

Mobile sensing technologies such as GPS, RFID, GSM, and Bluetooth and location-aware devices enabled for them have revolutionised the possibility to track moving entities such as vehicles, pedestrians, event visitors and animals. This tracking produces data about the movement of tracked entities commonly known as movement data. The increasing availability of tracking devices with a finer temporal resolution is leading to increasing availability of large volumes of movement data.

Movement data are an important resource in various application domains. For instance, understanding the transportation needs of urban dwellers as shown by their movement (e.g., commuting patterns (Dewulf et al., 2015)) is crucial for transportation management. In ecology, animal movement data are useful in studying different animal behaviours (e.g., migration (Cagnacci et al., 2011), foraging (Augustine and Derner, 2013) and mating (Long and Nelson, 2013)). In urban planning, the infrastructure development benefits from understanding the movement of the users of this infrastructure. Recreational park managers can benefit from profiling park users which can be achieved from their movement data (Meijles et al., 2014).

However, the usefulness of movement data in different application fields depends on the availability of tools and techniques for extracting human understandable information from these raw data. The field of computer science has a long tradition of developing tools and methods for data analysis. Special to mention is the Knowledge Discovery in Database (KDD), which is a process aimed at discovering knowledge from data (Fayyad et al., 1996). This process has been widely adopted for analysing huge classical data. Due to this feature of coping with large volume for data analysis, the KDD process has been adapted also for geographic data in the field of Geographic Information

Science (or GIScience). This special case of KDD has been named GKD (Geographic Knowledge Discovery) (Han and Miller, 2009). The Geographic Knowledge Discovery has also a potential for analysing movement data as a special type of geographic data characterised by a large volume and a temporal nature (Mennis and Guo, 2009).

The analysis of movement data has mainly focused on the extraction of movement patterns, which are detectable structures in the movement data of individuals or groups of individuals (Laube, 2009). However, patterns themselves are less useful in application fields because they include very limited or no high level description needed by the application fields. So, they need to be interpreted by relating them to the context in which the movement takes place. The context is important for interpreting movement patterns because the same movement pattern may have different interpretations in different contexts as explained in the following example illustrated in Figure 1.1.

The left part of Figure 1.1 shows a section of movement of taxis between San Francisco airport and the downtown. The figure shows that there exists a standard route on the map extent shown between these two places. The taxis normally follow the standard route as shown by many taxi positions recorded along this route (in brown). However, at some time a taxi follows an uncommon route (see the blue point positions). The interpretation of this movement pattern, called trajectory outlier, depends on the context. The right part of Figure 1.1 shows three examples of contexts that can lead to different interpretations of the trajectory outlier pattern. The first case suggests that the pattern may be interpreted as a deviation aimed at avoiding slow motion due to traffic congestion on the standard route. In another case (case 2), some roadwork may be in progress on the standard route; leading to intermittent blocking of the traffic. This may lead to a deviation aimed at avoiding "stopping and waiting" on the standard route. In a third case (case 3), a road on the standard route may be fully closed for some reason. In this case, the deviation will be aimed at replacing the unavailable standard route. These examples show that without considering the context it will be hard if not impossible to interpret the trajectory outlier pattern observed.

The literature contains some work relating movement patterns to specific examples of movement environment also called geographic context. For example, the movement of an individual has been related to the context made of other moving entities in (Laube, Imfeld and Weibel, 2005). In other examples, the movement of vehicles has been related to the road network to which the movement is restricted (Wang et al., 2014) while the movement of ships has been related to physical properties of the geographic space in which they travel (Lundblad et al., 2009). Beyond this integration of context examples, few authors attempted to develop a general framework to relate the movement to its

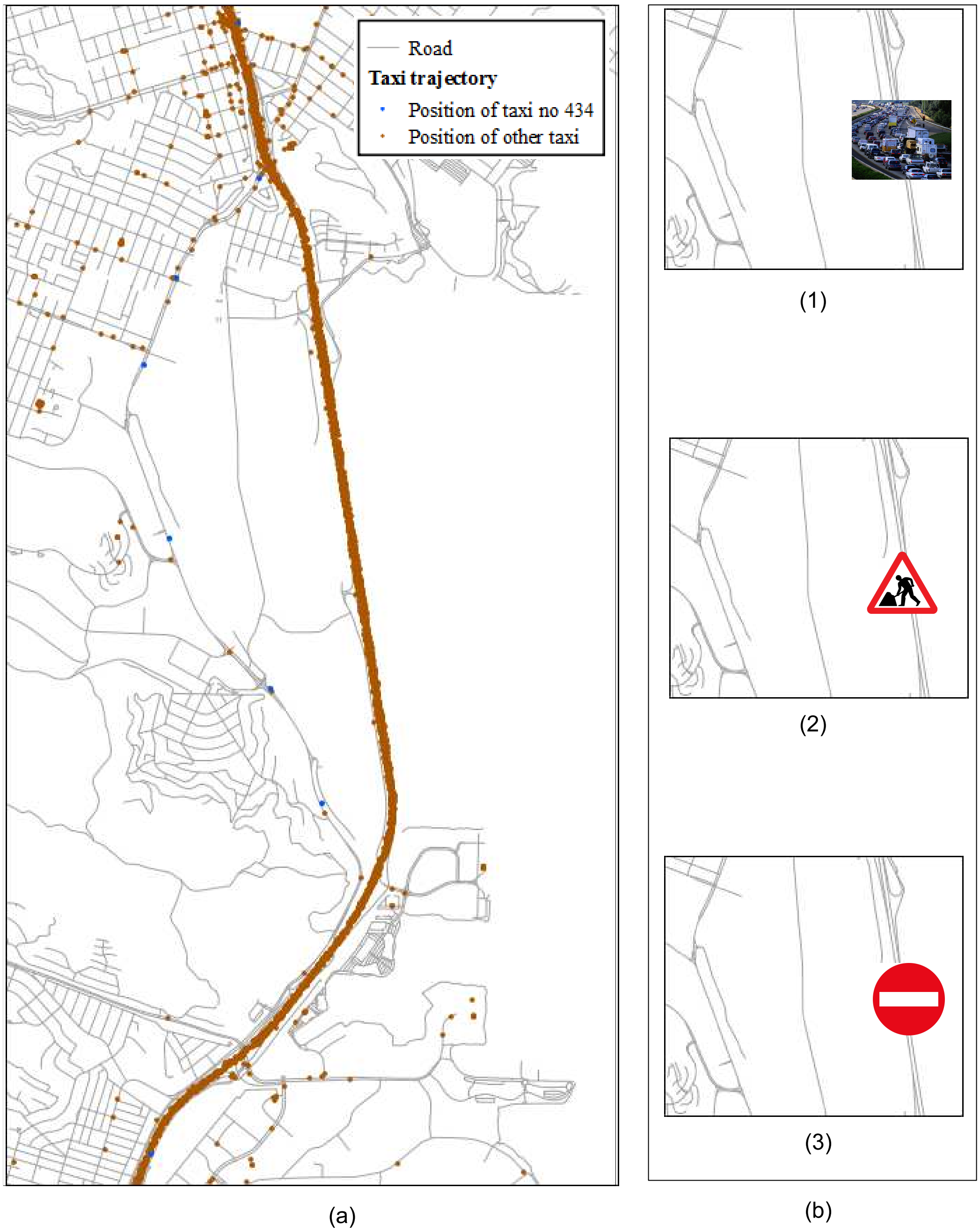



Fig. 1.1 A trajectory outlier pattern (a) and different contexts (b) that can lead to different interpretations for the pattern

environment in general considering the different types of this environment (Buchin et al., 2014; Andrienko, Andrienko and Heurich, 2011; Gschwend, 2015). However, a changing environment has not been considered in this general framework. The changing environment is hereafter referred to as a dynamic geographic context. For example, a street protest is a dynamic geographic context because it is associated with changes over time: participants' arrival, protest progress, and participants' dismissal. Other moving objects (e.g., pedestrians and other vehicles) constitute a dynamic context because their locations change over time. Likewise, a road segment considered along with its traffic congestion level is a dynamic context because the congestion level changes with time.

A dynamic geographic context affects the movement of objects depending on its dynamics. For instance, a street protest slows down the vehicles using the street or causes them to deviate from this street. Similarly, when the traffic congestion on a road segment increases the vehicles moving on it are forced to adopt a single file movement pattern and to move slowly. In a flock of migrating birds, the movement of an individual bird depends on the movement the other members of the flock: it has to change its direction to adjust it to the general direction of the flock. The geographic context that has been commonly considered in movement data analysis (e.g., points of interest (POIs)) is static, which simplifies relating it to movement. While checking the spatial proximity between car trajectories and a road intersection can allow identifying the cars that passed the intersection, it is not straightforward to identify cars that avoided a specific road segment because of a traffic jam on it. The reason is that unlike in the case of a static geographic context, in the case of a dynamic geographic context some relations between the movement and the context build over time, and a relation may be made of many different basic relations. The case illustrated in Figure 1.1 is used in the following as an example to explain these characteristics of a dynamic geographic context.

If we detect a traffic jam on the standard route, we can say that the taxi that deviated from the route (see the point positions shown in blue) had a "bypassing" relation with the traffic jam. This relation cannot be discovered from a snapshot of the positions of the taxi and the traffic jam; it builds over a time interval. Furthermore, this relation can be seen as a composition of other basic relations (example: approaching, keeping a distance, and moving away). The geographic context may expand over a wide space such that it presents spatial dynamics in addition to temporal dynamics. In the case of a traffic jam in Figure 1.1, the level of congestion at different locations along the route may be different and change with time. These peculiarities call for an approach which takes the time dimension into account explicitly to relate the movement to its embedding dynamic geographic context. Therefore, this thesis aims at integrating into the analysis of movement the embedding dynamic geographic context to support understanding of movement patterns. Relating movement to its embedding dynamic geographic context will help understand where, when and why the objects move the way they do. Though in some cases a moving entity may be wide and have some extent, throughout this thesis it is considered to be a point object.

## 1.2 Research objectives

The main objective of this thesis is to develop a conceptual and methodological framework for integrating a dynamic geographic context into the analysis of movement data. The aim in integrating the geographic context into the analysis of movement data is twofold: to understand the movement based on the context and to learn about the context from the movement. As an object moves it interacts with the environment in which it moves. With this consideration the thesis is guided by the following hypothesis: *Identifying and analysing interactions between a moving object and the dynamic geographic context supports understanding of the movement patterns and the movement context.* To achieve the objective, the following research questions are formulated:

- **RQ 1**: How can we model a dynamic geographic context so as to allow relating it to the movement it embeds?
- **RQ 2**: What are the sources of geographic context data for movement analysis and at which analysis step should these data be integrated?
- **RQ 3**: How can the relation between the movement and its embedding dynamic geographic context be explored in space and time to support movement pattern interpretation?
- **RQ 4**: How can the relation between the movement and its embedding dynamic geographic context be explored in space and time to support a deeper understanding of the geographic context?

## 1.3 Organisation of the thesis

This thesis consists of seven chapters as shown in Figure 1.2. After this introductory chapter, which presents the motivation, objectives and the structure of the thesis, subsequent chapters are organised as follows:

Chapter 2 presents important concepts used in the thesis and a review of related work. The chapter begins with an introduction to geographic data and spatial relations followed by an introduction to movement data. Then, the chapter presents a review on movement data analysis starting with an introduction of the KDD process and proceeding with different data mining methods for movement data. The last sections of the chapter are dedicated to the analysis of movement data taking into account the context.

Chapter 3 discusses an approach for modelling the dynamic geographic context of movement and relating this context to the movement it embeds. Different types of context elements and a model of interactions for linking them to movement are discussed.

Chapter 4 presents an analysis framework based on the KDD process for exploiting the relation between movement and its embedding geographic context to discover and interpret movement patterns, and further describe the context.

Chapter 5 presents an evaluation of the analysis framework proposed in chapter 4 through three experiments using real world data. The first experiment focuses on the temporal dynamics of the context while the second experiment focusses on the spatial dynamics. The third experiment uses a different form of context data and integrates these data at a different analysis step for comparison with the first two experiments. The work presented in this chapter is based on the papers published as contribution of this thesis.

Chapter 6 discusses the results with respect to the research questions that guided the thesis. This chapter also discusses the limitation of the work done in the thesis.

Chapter 7 concludes and gives the direction for future work.

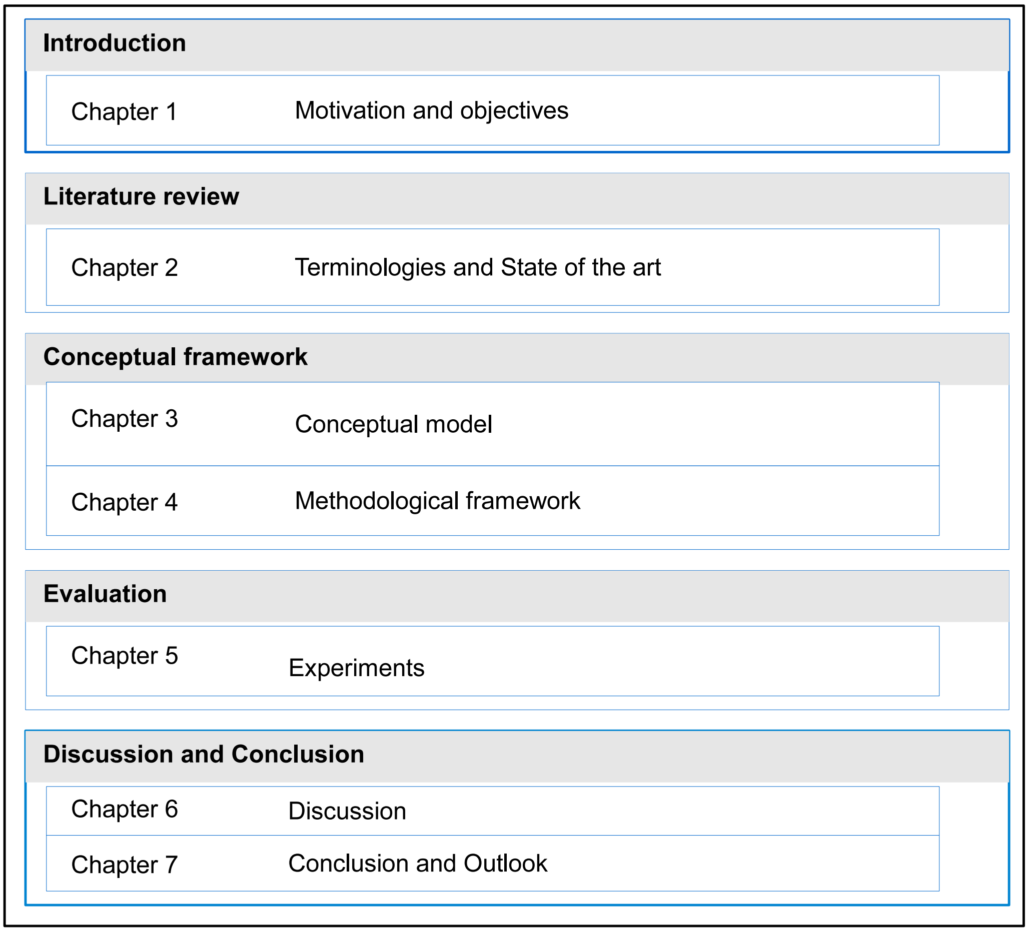



Fig. 1.2 Thesis structure

# Chapter 2

# Literature review

This chapter reviews the literature related to the main themes dealt with in the thesis. It is intended to introduce important concepts and provide the state-of-the-art on related work. The chapter starts with a brief introduction of geographic data and spatial relations, followed by an overview of movement data as a special type of geographic data. After this overview, which presents the nature and characteristics of movement data, the analysis of these data following the Knowledge Discovery (KD) process is presented. The last section of the chapter presents the state-of-the-art on integrating external data into the analysis of movement data. The last section includes also a survey of the sources of these external data.

## 2.1 Geographic data and spatial relations

### 2.1.1 Geographic data

**A geographic phenomenon** is defined as "*something of interest that can be named or described, georeferenced, and assigned a time (interval) at which it is/was present*" (de By, 2011). A geographic phenomenon for which, at every point in the study area a value can be determined, is called a field. Temperature and elevation are examples of a continuous field; that is, a field for which the value changes gradually from location to location. Land use and soil classifications are examples of a discrete field; that is, a field for which there is an abrupt change of value at some location. A geographic phenomenon which is not present everywhere in the study area is called an object. Examples of objects are hospitals and roads. A geographic phenomenon exists in real world. In order to study it using a computing system, a computer representation must be produced for it.

**Geographic data** (also known as spatial data) are data produced as computer representations of geographic phenomena. The first step in producing a computer representation of geographic phenomenon is to choose a model of the phenomenon. There are object-based models (points, polylines and polygons) and the field-based model (de By, 2011; Rigaux et al., 2001). Once a model for the geographic phenomenon has been chosen, a data structure is also chosen to store corresponding geographic data. There are two basic data structures to store geographic data: vector and raster (de By, 2011; Rigaux et al., 2001). A raster data structure is a grid with cells of equal size. Each cell is associated with a single value of a given attribute and the size of the cells defines the level of resolution. As a result, all variations within a cell are lost. The vector data structure is made of a finite number of points. A vector is made of one geometry attribute specifying the location and shape of the object, and other attributes containing other descriptions. In the vector data structure, a point is specified by its coordinates whereas lines and areas are usually represented by a sequence of points that are connected by straight lines.

### 2.1.2 Spatial relations

The analysis of geographic data focuses especially on spatial information. Spatial information includes the locations, shapes and sizes of geographic objects but also spatial relations between them. Spatial relations are classified into three categories: topological, directional, and distance relations.

**Topological relations** are relations that do not change under topological transformations such as rotation, scaling and translation (Rigaux et al., 2001). Topological relations have been widely studied. One of the important results of these studies is a formal model developed by Egenhofer (1991). The model, called *9-intersection model*, uses three primitives: *interior, boundary*, and *exterior* to define all possible topological relations. The object $\mathtt{A}$ being a point set, its *interior* is denoted by $\mathtt{A}^0$ while its *boundary* is denoted by $\partial A$. The *exterior* of $\mathtt{A}$, defined as its complement, is denoted by $\mathtt{A}^-$. The topological relationship between two objects $\mathtt{A}$ and $\mathtt{B}$ is then described by the 9-intersection matrix $\mathtt{I}$ resulting from intersecting the elements in their corresponding sets {$A^0$, $\partial A$, $A^-$} and {$B^0$, $\partial B$, $B^-$}:

$$I(A,B) = \begin{pmatrix} \partial A \cap \partial B & \partial A \cap B^0 & \partial A \cap B^- \\ A^0 \cap \partial B & A^0 \cap B^0 & A^0 \cap B^- \\ A^- \cap \partial B & A^- \cap B^0 & A^- \cap B^- \end{pmatrix}$$

Based on the 9-Intersection model, there are eight possible topological relations between two 2-dimensional areas with connected boundaries. Figure 2.1 shows these relations together with their corresponding intersection matrices. Extended models have been developed to describe relations between objects of different dimensions (e.g., an area and a line (Clementini et al., 1993)) and complex areas with holes (Egenhofer et al., 1994).

**Directional relations**. A direction is commonly defined for a line segment to indicate its angle with respect to a given fixed direction in the frame of reference, typically geographic north or the positive x or y axis (Frank, 1992). The direction can be represented by a numeric function for which values are real numbers (e.g., in degrees) or qualitatively by predicates based on the four cardinal directions (`north, east, south, west`). Frank (1992) proposed an algebraic approach for reasoning on directional spatial relations qualitatively.

A qualitative direction is a function that maps two points or a line segment onto a symbolic direction. The `n` different symbols available for describing directions make a set $C_n$ and depend on the specific system of directions used. For example, $C_4$={N, E, S, W} where the symbols stand for `North, East, South,` and `West` respectively. Papadias and Theodoridis (1997) proposed a generalisation of the directional relations from points to areas for a projection-based cardinal reference system. They applied the definitions of the relations between points using universally and existentially quantified formulae on points of the areas. The quantified formulae allowed differentiating more relations using the qualifiers `strong, weak` and `just`. For instance, the relation `strong_north(P,Q)`means that all points of `P` are north to all points of `Q`.

**Distance relations**. A distance relation is represented by a function which maps from a pair of points to a positive real number. The positive real number expresses the distance in a suitable unit (e.g., meters). However, a distance can also be represented qualitatively using predicates such as far and close. The predicates correspond to symbols from a finite set such as `D2={C, F}` for ‘`close`’ and ‘`far`’ or `D4 = {CC, C, M, F, FF}` for ‘`very close`’, ‘`close`’, ‘`medium`’, ‘`far`’, and ‘`very far`’.

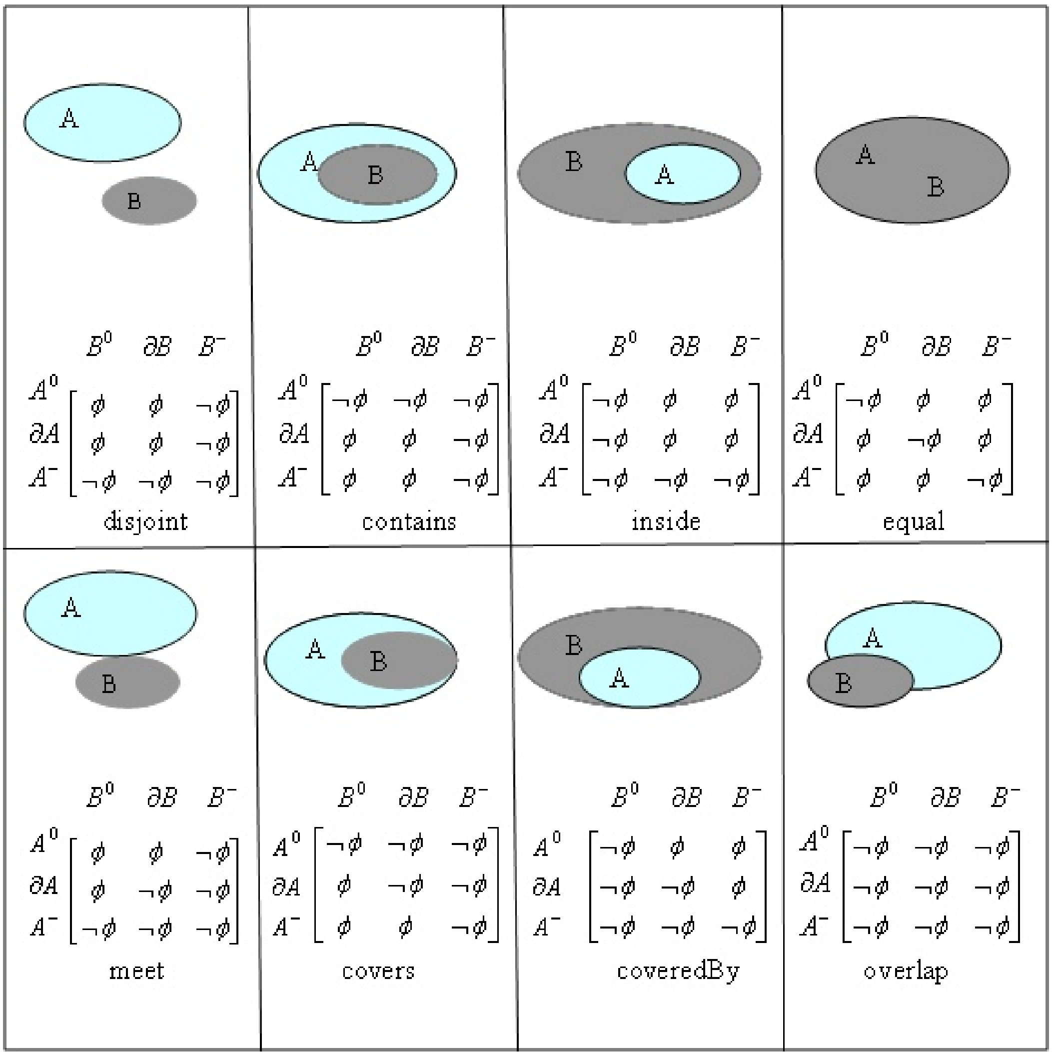
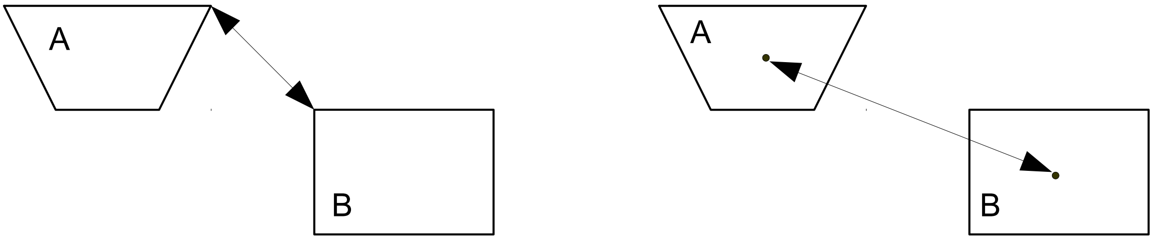



Fig. 2.1 The eight topological relations between two regions with connected boundaries (based on Egenhofer (1991))

*Euclidean distance.* The Euclidean distance between two points $a, b \in R^n$ is given by:

$$d(a,b) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2}$$

In case of lines and areas, the Euclidean distance is considered to be the shortest distance between two points on either object. For the case of areas, the distance between the centroids of the objects can also be used (see Figure 2.2). In case there are obstacles between two objects the distance between them is often calculated as the “travel distance”. The “travel distance” is the length of the shortest path in the




(a) $\boldsymbol{d(A,B) = \min(d(a,b)|a \in A, b \in B)}$ (b) $\boldsymbol{d(A,B) = d(centroid(A), centroid(B))}$

Fig. 2.2 Distance between two areas: minimal distance (a) and centroid distance (b)

road network between two locations. For example, the distance between two buildings located on both sides of a river is the distance along the shortest available path through a bridge. Euclidean distance measure is not suitable for long distances on the surface of the Earth due to the shape of the Earth. In this case, the Great-circle distance is more suitable.

*Great-circle distance on Earth.* The great circle distance (or spherical distance) is defined as "the shortest distance between points x and y on the surface of the Earth measured along a path on the Earth's surface" (Deza and Deza, 2009). It is the length of the great circle arc, passing through `x` and `y`, in the spherical model of the planet. Given the latitude and longitude of `x` ($\delta_1$ and $\varphi_1$respectively) and the radius of the Earth `r`, the great-circle distance between `x` and `y` is equal to

$$r \arccos(\sin\delta_1 \sin\delta_2 + \cos\delta_1 \cos\delta_2 \cos(\varphi_1 - \varphi_2))$$

### 2.1.3 Qualitative spatial representation and reasoning

Qualitative spatial relations are very common. However, since they are essentially on a nominal scale they provide limited capabilities for making comparisons and analyses beyond an exact matching. To overcome this limitation two models are widely used: the *composition table* and the *Conceptual Neighbourhood graph (CNG)*. These models especially support qualitative reasoning. Qualitative reasoning is about deducing new knowledge from existing qualitative knowledge. In qualitative reasoning the existing knowledge is expressed by variables which can only take a small predetermined number of values and the inference rules use these values and not numerical quantities approximating them (Frank, 1991).

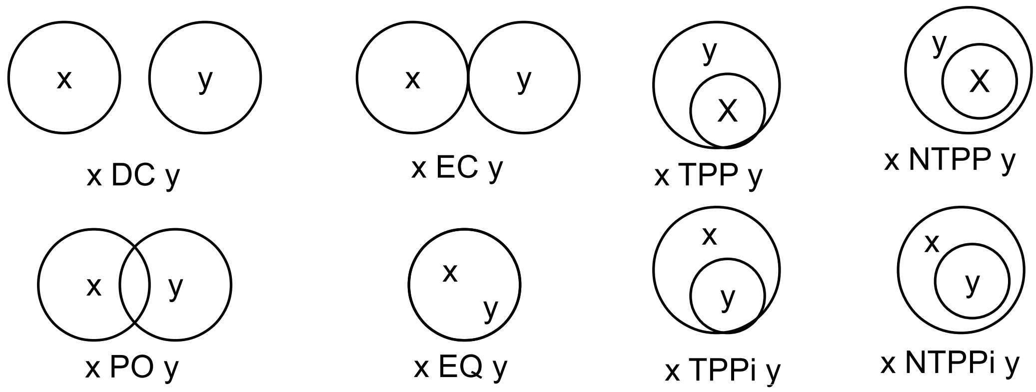



Fig. 2.3 RCC-8 relations (based on Randell et al. (1992))

In a composition table different possible relations are represented as columns and as rows. The intersection cell between a row and a column contains the relations that can be inferred when the relations at the row and the column hold. The composition table allows making a compositional inference; that is, from two relational facts of the form $R_1$`(a,b)` and $R_2$`(b,c)`, deducing a relational fact of the form $R_3$`(a,c)` involving only `a` and `c` (Cohn and Hazarika, 2001). For example, the region connection calculus (RCC) describes regions by their possible relations to each other. It consists of eight basic relations (see Figure 2.3): *disconnected (DC), externally connected (EC), equal (EQ), partially overlapping (PO), tangential proper part (TPP), tangential proper part inverse (TPPi), non-tangential proper part (NTPP), non-tangential proper part inverse (NTPPi).* The composition table of these relations is shown in Table 2.1. This composition table allows the following inference for example. If `Region a` is Tangential Proper Part (TPP) of `Region b` and `Region b` is Tangential Proper Part (TPP) of `Region c` then `Region a` is either Tangential Proper Part of `Region c` or Non-Tangential Proper Part of `Region c`.

The idea of conceptual neighbourhood graphs has been introduced by Freksa (1991) on interval relations where he linked two interval relations as conceptual neighbours if a smooth transformation can be performed between them. Freksa (1991) states that *"two relations between pairs of events are conceptual neighbours if they can be directly transformed into one another by continuous deformation (i.e., shortening or lengthening) of the events"*. After their introduction on interval relations (Freksa, 1991), conceptual neighbourhood graphs have been used on other types of qualitative relations such as topological relations (Egenhofer and Mark, 1995; Kurata and Egenhofer, 2006; Reis et al., 2008), lines of sight relations (Galton, 1994), relations between moving objects (Van de Weghe and De Maeyer, 2005), and line segment relations (Schlieder,

| $R_1(a,b)$ \ $R_2(b,c)$ | DC | EC | PO | TPP | NTPP | TPPi | NTPPi | EQ |
|---|---|---|---|---|---|---|---|---|
| **DC** | no info | DC, EC, PO, TPP, NTPP | DC, EC, PO, TPP, NTPP | DC, EC, PO, TPP, NTPP | DC, EC, PO, TPP, NTPP | DC | DC | DC |
| **EC** | DC, EC, PO, TPP, NTPPi | DC, EC, PO, TPP, TPPi, EQ | DC, EC, PO, TPP, NTPP | EC, PO, TPP, NTPP | PO, TPP, NTPP | DC, EC | DC | EC |
| **PO** | DC, EC, PO, TPP, NTPPi | DC, EC, PO, TPP, NTPPi | no info | PO, TPP, NTPP | PO, TPP, NTPP | DC, EC, PO, TPPi, NTPPi | DC, EC, PO, TPP, NTPPi | PO |
| **TPP** | DC | DC, EC | DC, EC, PO, TPP, NTPP | TPP, NTPP | NTPP | DC, EC, PO, TPP, TPPi, EQ | DC, EC, PO, TPP, NTPPi | TPP |
| **NTPP** | DC | DC | DC, EC, PO, TPP, NTPP | NTPP | NTPP | DC, EC, PO, TPP, NTPP | no info | NTPP |
| **TPPi** | DC, EC, PO, TPP, NTPPi | EC, PO, TPPi, NTPPi | PO, TPPi, NTPPi | PO, TPP, TPPi, EQ | PO, TPP, NTPP | TPPi, NTPPi | NTPPi | TPPi |
| **NTPPi** | DC, EC, PO, TPP, NTPPi | PO, TPPi, NTPPi | PO, TPPi, NTPPi | PO, TPPi, NTPPi | PO, TPP, NTPP, TPPi, NTPPi, EQ | NTPPi | NTPPi | NTPPi |
| **EQ** | DC | EC | PO | TPP | NTPP | TPPi | NTPPi | EQ |

Table 2.1 The composition table for RCC-8 relations (based on Cohn et al. (1997))

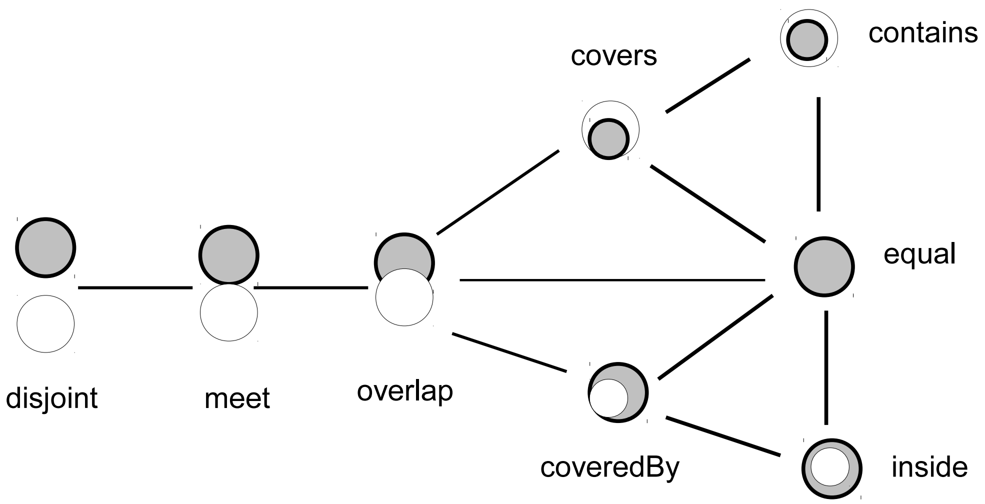



Fig. 2.4 The Conceptual neighbourhood graph of topological relations between regions with connected boundaries (based on Egenhofer and Al-Taha (1992))

1995). Figure 2.4 shows a conceptual neighbourhood graph of the topological relations between two regions with connected boundaries (see Figure 2.1). The topological relations *meet* and *overlap* are conceptual neighbours because a gradual change can transform the *meet* relation into overlap relation without any other intermediate relation. In contrast, the relations *meet* and *covers* are not conceptual neighbours because a gradual change of the *meet* relation transforms it firstly into another relation, namely *overlap*.

There are two approaches commonly used for deriving a conceptual neighbourhood graph: the *snapshot* approach and the *smooth transitions* approach (Egenhofer and Mark, 1995). The snapshot approach derives the neighbourhoods by comparing two different relations without any knowledge about the potential transformations that may have caused the change from one to the other. Neighbours are selected based on least noticeable differences as indicated by a conceptual distance between the relations (e.g., a topological distance computed on corresponding 9-intersection matrices). The smooth transition approach develops neighbourhoods based on the knowledge of the deformations that may change a relation (for example: knowing that one objects moves towards the other).

## 2.2 Movement data

Movement is change of position of an entity over time. The movement of different types of entities such as people, vehicles, animals, and natural phenomena (e.g., hurricanes and tornadoes) may be of interest and hence recorded. The data representing the movement of an entity are commonly referred to as movement data. There are other terms that have been used to refer to these data such as trace data or traces (Pan et al., 2013), lifeline (Laube, van Kreveld and Imfeld, 2005), mobility data (Pelekis and Theodoridis, 2014), and trajectory data or trajectories (Zheng and Zhou, 2011). Although some authors (Spaccapietra et al., 2013) tried to differentiate between these terms, they are sometimes used interchangeably. The common concept behind them is a temporally ordered sequence of positions that the moving entity took during its movement.

Movement is inherently continuous, but practical limitations dictate that movement data are discrete; that is, only some positions of the moving entity are recorded. Since each sampled position has a location and a time moment at which it is recorded, movement data are spatio-temporal in nature. In its basic form, the trajectory of a moving object can be formally represented as: $\mathtt{T}=<\mathtt{p}_1 \ \ldots \ \mathtt{p}_n>$ where $\mathtt{p}_k=(\mathtt{id}_k,\mathtt{loc}_k,\mathtt{t}_k)$ is the $\mathtt{k}^{\mathtt{th}}$ position, $\mathtt{id}_k$ is the position identifier, $\mathtt{loc}_k$ is the spatial location of the position, and $\mathtt{t}_k$ is the time at which the position was recorded. Figure 2.5 shows an example of movement data. The figure shows a trajectory of one taxi moving in the city of Rome. The trajectory is a temporally ordered sequence of points representing some positions that the taxi occupied. The points are connected with line segments to construct the trajectory as a polyline. The order of the points is indicated by their timestamps as shown in the magnified small section of the trajectory in Figure 2.5. As it can be seen from Figure 2.5, the distance and time interval between successive sampled positions may be different.

The advances in sensor and communication technologies have led to different forms of movement data depending on the technology used for recording the movement. Spinsanti et al. (2013) differentiated *GPS* (Global Positioning System), *GSM* (Global System for Mobile Communications), and *Geo-social network* based trajectory data. Pelekis and Theodoridis (2014) differentiated two other forms; *RFID* (Radio Frequency Identification) based and *Wi-Fi* based data. GPS based data are a temporally ordered sequence of geographic coordinates recorded by a GPS-enabled device carried by the moving entity. GSM based data are a temporally ordered sequence of identifiers of the cells through which the moving object passed. Geo-social network based data are content found on Internet social media websites to which geographic coordinates have

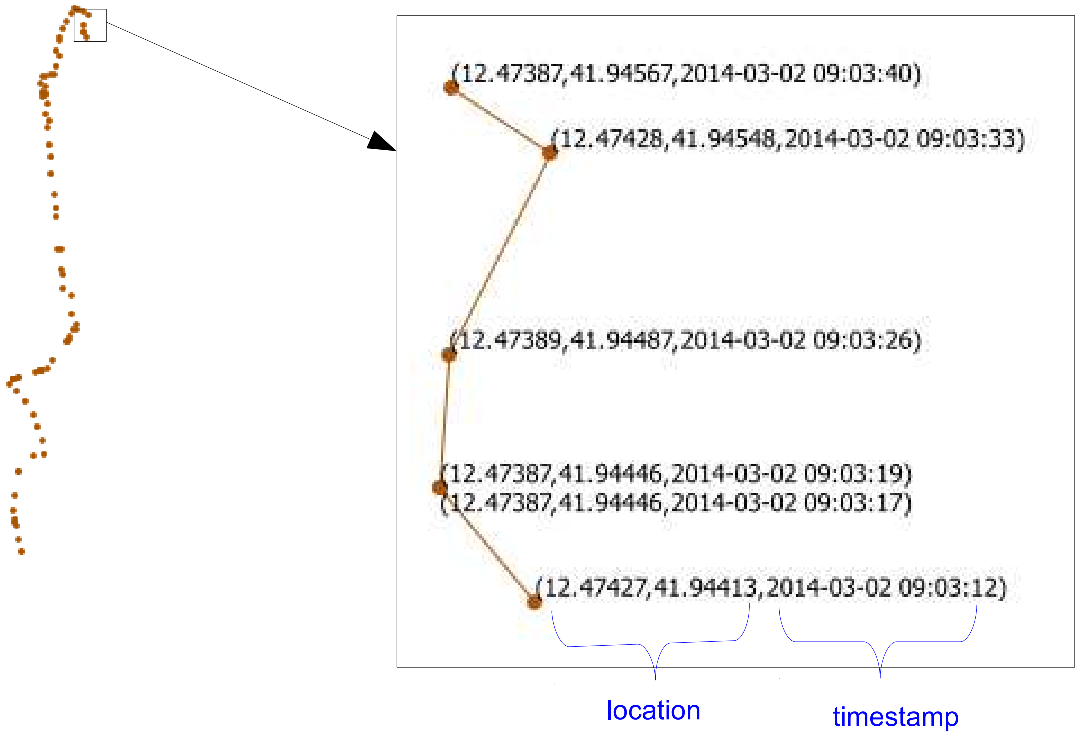



Fig. 2.5 Example of movement data

been attached. RFID based data are made of a sequence of identifiers of RFID readers through which the moving object passed, while Wi-Fi based data are a sequence of identifiers of access points that communicated with the moving object.

The localisation technologies commonly used for recording movement are compared in Table 2.2. The "accuracy" indicates how correct the location recorded is. For example, from Table 2.2 it can be noted that the location of a moving entity may be recorded up to 20 meters away from its actual position if the movement is recorded using the WiFi technology. The "coverage" indicates the distance range (with respect to reference object if any) in which the position of the moving entity can be recorded. For example, with GPS technology the position can be recorded anywhere in outdoor while with WiFi technology it can be recorded only in less than 100 meters of the access point.

Movement data are inherently uncertain because of different types of impreciseness. For example, the inaccuracy of the technology used for movement recording (see Table 2.2) leads to trajectory points located away from the actual positions they represent. Likewise, if the location is manually entered human errors may cause incorrect point location with respect to the actual position. The uncertainty of

| Technology | Location data | Accuracy | Coverage |
|---|---|---|---|
| **GPS** | Geographic coordinates | 1 – 5 meters | outdoors |
| **WiFi** | Access point ID | 1 – 20 meters | < 100 meters from an access point |
| **GSM** | Cell tower ID | 50 – 200 meters in cities | Cell coverage. 5 – 30 km from a cell tower |
| **Bluetooth** | Device ID | Sensing range of Bluetooth | 5 – 10 meter for Class 1; 20 – 30 meters for Class 2 |
| **RFID** | Reader's ID/Position | Sensing range of RFID | 1 meter for passive RFID; 100 meters for active RFID |

Table 2.2 Comparison of common localisation technologies for tracking moving entities (source: Pan et al. (2013))

movement data can also be caused by the sampling frequency adopted for movement recording. If the sampling frequency is low the intermediate positions of the moving entity will be highly uncertain. The uncertainty due to low sampling frequency is especially observed on GSM based and geo-social network based movement data. For GSM based data, a position is recorded when the mobile phone interacts with a cell tower (e.g., a call is made or received, a message is sent or received), which may happen after a long time interval. For geo-social network based data, a position may be recorded only at places selected by the user (e.g., when the user arrives at specific places), which may happen with a time interval of several days. The inherent uncertainty of movement data makes it necessary to pre-process the data for sanitising them before the actual analysis.

## 2.3 Movement data analysis

Movement data are voluminous in their nature due to a high rate and a long duration of movement recording. Because of this voluminous nature a common approach of analysing movement data follows the *Knowledge Discovery* (or *Knowledge Discovery in Database (KDD)*) process (Fayyad et al., 1996; Maimon and Rokach, 2010), which is designed to extract useful information from huge datasets.

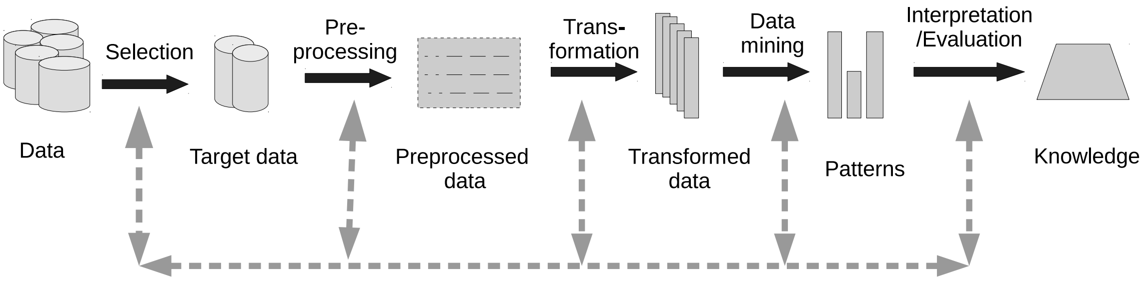



Fig. 2.6 The Knowledge Discovery in Database (adapted from Fayyad et al. (1996))

### 2.3.1 Knowledge Discovery

The Knowledge Discovery is a process aimed at extracting useful information (knowledge) from a huge dataset. As shown in Figure 2.6, the process is iterative which means that at any step in the process, it is possible to loop back to some previous step.

The Knowledge Discovery process starts with a preparation of data to be analysed. Firstly, a subset of data from which knowledge is to be discovered is selected. Next, the selected subset is pre-processed by carrying out operations aimed at sanitising the data. The pre-processed data are then transformed to adapt them to the requirements of the analysis method. At the core of the knowledge discovery process is the data mining step. *Data mining* is the essential step where specific algorithms are applied on the data to extract patterns Kamber et al. (2012). A pattern is an interesting structure found in the data. When patterns are interpreted and evaluated, useful information (also called knowledge) is obtained. Since the steps before "Data mining" comprise preliminary operations that prepare the data for actual analysis they can be considered together as one step called pre-processing.

### 2.3.2 Knowledge Discovery in movement data

The overall aim of analysing movement data is to extract and understand movement patterns for some application. Based on the Knowledge Discovery process (Figure 2.6), the analysis of movement data starts by selecting the part of data to be analysed. The selected data are then pre-processed. After the pre-processing, one or more data mining methods are applied on the pre-processed data to discover patterns, which are finally interpreted and evaluated.

### Movement data pre-processing

The pre-processing of movement data is an important step especially because these data are often not accurate. The pre-processing step includes, among others, tasks such as data cleaning, data compression, map matching, and trajectory segmentation (Zheng, 2015). Data cleaning aims at filtering noise from the data. For example, GPS points that fall away from the study area due to signal recording errors are discarded. Trajectory compression aims at reducing the amount of sample points to reduce the volume of data to be processed. For example, in case the density of recorded GPS points is not important several closely located points can be replaced by one representative point. Map matching aims at matching trajectories to the road network (Brakatsoulas et al., 2005). This is done in case the movement follows a road network to ensure that the next steps use the right position of trajectory points with respect to the road network. Trajectory segmentation divides a trajectory into meaningful sub-trajectories required for subsequent operations.

### Movement data mining

Data mining methods commonly applied on classical data have also been applied on movement data. In this direction, clustering and classification methods are either used alone or in combination with other analysis methods (e.g., statistical methods) to form more complex mining methods.

#### *Clustering*

Movement data clustering aims at grouping trajectories into a finite set of categories, also called clusters, based on their characteristics. The trajectories in the same cluster exhibit movement characteristics that are similar and different from those of trajectories in other clusters. State-of-the-art clustering algorithms for trajectories are extensions of traditional clustering algorithms through a proper definition of distance (or similarity) functions. The distance (or similarity) function is used to determine which trajectories belong to the same cluster and its choice depends on the application (Rokach, 2010). The similarity function can be, for example, having same route, same destination, same origin, same route and destination, or same direction. Among traditional algorithms that have been extended, two well-known algorithms can be mentioned: DBSCAN (Density Based Spatial Clustering of Applications with Noise) (Ester et al., 1996) and OPTICS (Ordering Points To Identify the Clustering Structure) (Ankerst et al., 1999). As examples of such extensions, T-OPTICS (Nanni and Pedreschi, 2006) was developed as an extension of OPTICS by defining a spatio-temporal distance for comparing and clustering trajectories while ST-DBSCAN (Spatial-Temporal DBSCAN) (Birant and Kut, 2007) uses two parameters for similarity measure to improve the identification of clusters and noise.

Trajectory clustering can be applied on either whole trajectories or sections of trajectories depending on the goal of analysis and the similarity function applied. For example, with "similar origin and similar destinations" as the similarity function, the route followed is not important and therefore, the trajectories can be clustered as whole. An example of this case can be seen in (Gaffney et al., 2007) where the interest is on the overall directions of extra-tropical cyclone trajectories for obtaining clusters such as `south-to-north oriented` and `west-to-east oriented`. On the other hand, if the interest is on different locations traversed by the trajectories, for example the similarity is defined as having visited the same types of places; the clustering is applied on sections of trajectories and based on the clusters of sub-trajectories the cluster membership of the whole trajectory is decided. This case is exemplified by the TraClus clustering algorithm (Lee et al., 2007).

#### *Classification*

Trajectory classification aims at categorising trajectories into one of predefined categories based on some features of the trajectories. For example, the trajectory classification may be aimed at labelling each trajectory from a large set with its transportation mode, given a small set of trajectories assigned transportation modes.

Most trajectory classification algorithms follow a traditional train and test approach: they first extract a set of discriminative features and then use them to train an existing standard classification model, which is finally used to perform the classification. Among the standard classification models that have been commonly applied in trajectory classification there are decision trees and Support Vector Machine (SVM). For example, Zheng et al. (2010) used the Decision Tree to classify trajectories into different transportation modes based on average velocity of a segment, heading change rate, and velocity change rate as discriminative features. The same classification problem has been addressed by Bolbol et al. (2012) using SVM on the speed and acceleration.

#### *Pattern mining*

The objective in trajectory pattern mining is to discover and describe the movement patterns hidden in trajectories. The literature reports a large number of types of movement patterns, which have also been put into an integrated view by some authors

(Dodge et al., 2008). Likewise, several methods have been developed for mining these patterns. The methods can be put into three categories: *repetitive pattern mining, frequent pattern mining*, and *group pattern mining.*

*Repetitive pattern mining* concerns regular movement patterns such as the movement of a commuter, which is repeated every day, or the movement of a migratory bird, which is repeated every season. A common approach in discovering repetitive movement patterns, also called periodic movement patterns, is to apply the mining on sequences of regions. In earlier studies (Mamoulis et al., 2004; Cao et al., 2007), the period (i.e., the duration after which the movement returns to the same location) had to be specified for discovering the sequence of regions. Unlike the earlier studies, the *Periodica* algorithm (Li, Ding, Han, Kays and Nye, 2010) detects automatically the period.

*Frequent pattern mining* is about extracting (parts of) routes that have been frequently followed by the moving objects in the trajectory dataset. Frequent trajectory patterns can be defined using spatial or spatiotemporal characteristics of the trajectories. The definition based on spatial characteristics considers only the sequence of the locations visited (see Figure 2.7(a)). The *frequent spatiotemporal sequential patterns* (Cao et al., 2005) and the *Generalised Sequential Patterns* (Orellana et al., 2012) exemplify this case. The definition based on spatiotemporal characteristics considers, in addition to the sequence of the locations visited, the transition time between the locations. The *T-Patterns* (Giannotti et al., 2007), illustrated in Figure 2.6(b), are an example of this second case. The T-Patterns shown in Figure 2.7(b) mean that location B follows location A and the transition between these two locations takes between 10 and 12 minutes. A common approach to mining frequent patterns consists in finding important regions from the trajectories and then applying sequence mining on a temporally annotated sequence of these regions (Giannotti et al., 2007; Kang and Yong, 2010).

*Group pattern mining* aims at extracting movement patterns involving groups of objects that move together. The objects that form a group must stay close in space for a considerable period of time. Several group patterns have been defined based on the general condition of spatio-temporal closeness, the internal structure of the group and the characteristics of group members. The mostly studied group patterns are *flocks* (Benkert et al., 2008; Wachowicz et al., 2011), *convoy* (Jeung et al., 2008), and *swarm* (Li, Ding, Han and Kays, 2010). Figure 2.8 shows examples of these three patterns. A flock is a group of at least `m` objects that travel together for at least `k` consecutive timestamps such that at any of these timestamps they are found within a disc of radius

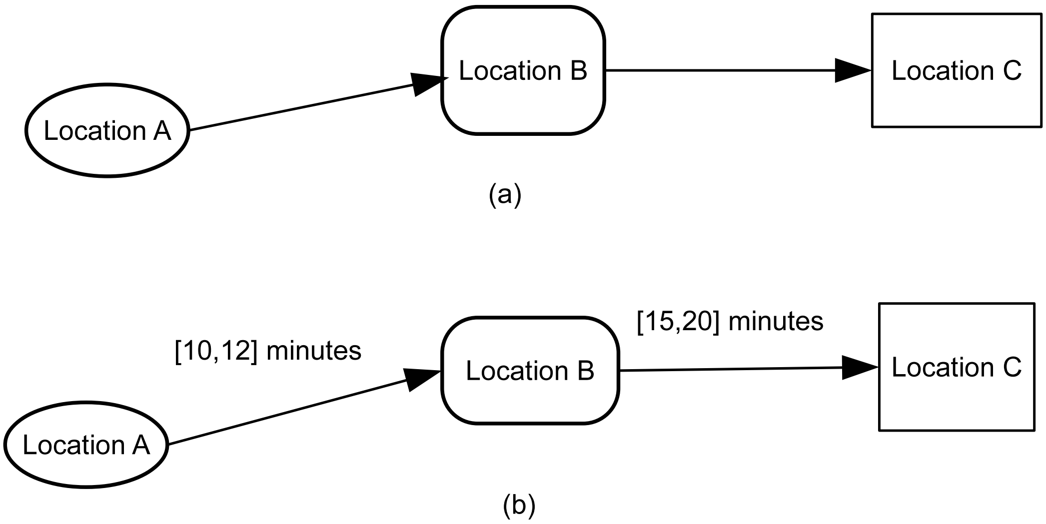



Fig. 2.7 Examples of frequent patterns: a) a spatiotemporal sequential pattern, and b) a T-pattern

`r`. The pattern is defined by the three parameters `m, r`, and `k`. The convoy pattern relaxes the disc shape constraint to represent a group of moving objects forming any shape. It is a group of at least `m` objects travelling together for at least `k` consecutive timestamps such that at each of these timestamps the group can be found using a density-based clustering with parameters `d` as the neighbourhood distance and `m` as the minimum number of objects. The swarm pattern is an extension of the convoy pattern, which further relaxes the constraint of consecutive timestamps. It is defined by the same parameters as the convoy except that in a swarm `k` is the minimum number of timestamps at which the group is found irrespective of whether the timestamps are consecutive or not. For example, with `m=4` and `k=2` the pattern in Figure 2.8 (c) is not a convoy because only three objects ($O_1$, $O_2$, and $O_3$) are found in the group for at least two consecutive timestamps. However, this pattern is a swarm because at two timestamps ($t_1$ and $t_3$) the group includes four objects, which is the minimum required and the swarm pattern does not require the timestamps to be consecutive.

The category of group patterns also includes patterns in which the members of the group have some interaction in which each member has a specific role. An example of such pattern is the *leadership* pattern (Andersson et al., 2008) in which the group comprises a leader which moves ahead and followers for the time duration of the pattern. Another example is the *chasing* pattern (de Lucca Siqueira and Bogorny, 2011) in which the object moving ahead aims at escaping the follower, which tries to reach it.

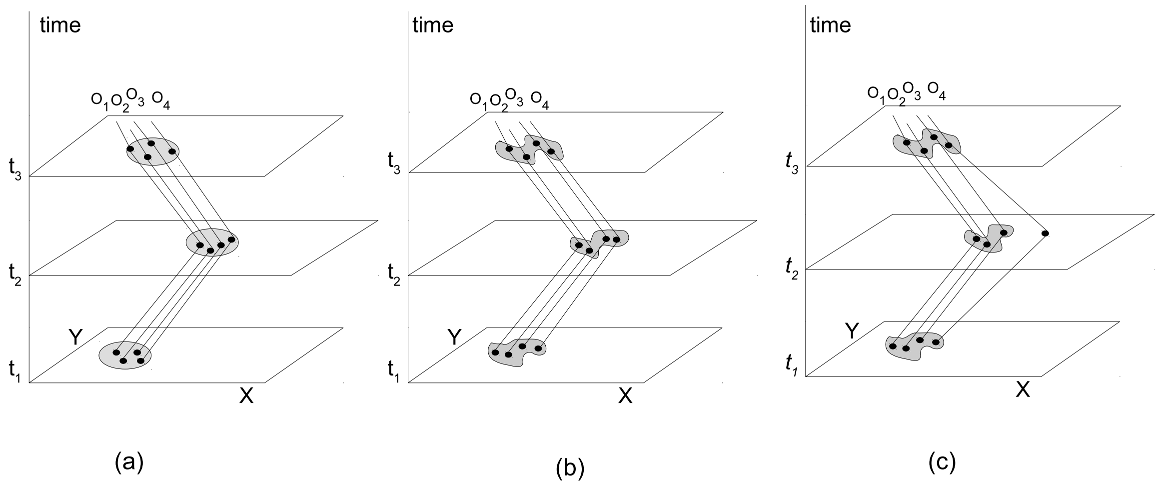



Fig. 2.8 Examples of group pattern: (a) Flock, (b) Convoy, and (c) swarm

Different examples of group patterns have been discussed under the general concept of *Relative Motion* (REMO) (Laube, Imfeld and Weibel, 2005).

The common approaches for group pattern mining involve clustering methods and checking the condition on parameters that define the pattern such as the minimum number of group members and the minimum duration of the pattern.

### Movement pattern interpretation/evaluation

The patterns mined in the previous step are further processed to turn them into useful information (also called *knowledge*). Visual analytics tools have been developed to support in interpreting and evaluating movement patterns. These tools support the analyst to get insights into the movement patterns by providing a convenient interface to explore trajectories and movement patterns. From this interface the analyst can filter trajectories and patterns based on such criteria as a time interval, a spatial window, and attribute values (Andrienko et al., 2008; Keler and Krisp, 2016). The analyst can also explore the patterns from different views (Guo et al., 2011). Different functionalities provided by visual analytics tools facilitate linking trajectories and movement patterns to analyst's prior knowledge and evidence from other sources. The interpretation of the observed movement patterns comes from reasoning about the patterns and the linked additional information. In order to support the analysis of large movement dataset, the basic visual analytics methods such as those presented in (Andrienko et al., 2008) are often supported by computational methods as done in (Andrienko, Andrienko, Hurter, Rinzivillo and Wrobel, 2011) and (Andrienko and Andrienko, 2013).

### 2.3.3 Spatio-temporal analysis

The KDD process discussed in previous sections originated from the field where it was applied on classical data in which the space and/or time dimensions are not present or may be ignored. However, as discussed in section 2.2 movement data are spatio-temporal in nature. In this section important methods that focus on combined handling of the space and time dimensions are presented. They are put in two categories: qualitative reasoning and visual analysis.

Van de Weghe (2004) developed the Qualitative Trajectory Calculus (QTC) for representing and reasoning qualitatively about movements of objects. Unlike the theories presented in previous sections for qualitative spatial representation and reasoning (e.g., see RCC-8 in section 2.1.3) QTC embeds consideration of the time dimension in addition to the space dimension. Depending on the level of details, different types of QTC are defined to describe the relationships between a pair of moving objects. The basic form of QTC compares positions of objects at different time points. Based on how the distance between the objects changes, QTC uses three symbols (+, -, 0) to represent that an object moves away from or towards the other, or is stable with respect to the other respectively. Other forms of QTC use more labels to represent more details about the relationship between the objects. For example, a relationship represented by (+ - 0) means that the first object is moving away from the second (+), while the second object is approaching the first (-), and both objects have the same speed. In that relationship, the first label represents the movement of the first object with respect to the second object. The second label represents the movement of the second object with respect to the position of the first object while the third label represents the relative speed of the first object with respect to the second. An example of qualitative reasoning about space and time using QTC is presented in (Van de Weghe et al., 2006).

In the area of visual analysis of movement data, the space-time cube (Kraak, 2003) is commonly used. It represents the geographic space with the X-Y plane and the time with the Z-axis. Trajectories of moving objects are then shown as 3D polylines in the cube. Tominski et al. (2012) developed a variant of the space-time cube called “trajectory wall”, which stacks 3D color-coded bands on a 2D map. The major problem of the space-time cube is the visual clutter that is observed in case of a large number of trajectories. In order to address this problem, different methods have been developed to aggregate trajectories and visualise their aggregate. For example, Demšar and Virrantaus (2010) introduced the concept of *space–time density of trajectories*. This concept generalises the standard 2D kernel density around 2D point data into 3D density around 3D polyline data (i.e., trajectories). The density around each trajectory in space and time is calculated as a volume such that the value is assigned to each voxel according to the distance of the central point of the respective voxel to the trajectory. By adopting a different way of calculating kernels and based on the *space–time density of trajectories* method, Demšar et al. (2015) developed a new method called *stacked space-time densities of trajectories*. In the new method, kernels are calculated per voxel layer producing probability layers which are then stacked one upon another to form the stacked space-time density volume.

With the same idea of aggregating movement into densities, Krisp et al. (2013) adopted a different approach which visualises moving 2D densities. This approach, called *directed kernel density estimation* (DKDE) initially introduced in (Krisp and Peters, 2011), considers the temporal dimension through the direction of movement. For each moving object, a movement vector is computed from the point positions at two different times. Density values are calculated using a kernel function defined as a 2D projection of a cone constructed on the movement vector based on the speed and direction of movement. Another approach used to alleviate the visual clutter problem involves linking to the space time cube multiple displays which portray different information for visual analysis (Zhao et al., 2008).

## 2.4 Movement data analysis and context data

Movement data in their basic form, as presented in section 2.2, lack semantic information that would support their analysis, especially the pattern interpretation/evaluation phase. In order to improve the analysis in general and the pattern interpretation in particular, analysts integrate into the analysis other related data that we call context data. This section introduces context data for movement and presents the state-of-the art in incorporating them into movement data analysis.

### 2.4.1 Context data

The term "context" is used in many different applications with a meaning of situation. For example, user information, device information and environment information are used as context data for generating the appropriate interface in user interface design (Hariri et al., 2008). In information retrieval, information on user's on-going activity is used as context data for retrieving information which is most relevant for the user (Vallet et al., 2007). Likewise, in data repository management, information about users such as their locations and roles in the application concerned is used to tailor response to data access queries (Bolchini et al., 2011). In a pervasive environment with a lot of services, information such as location, time and life status of a user is used as context data for selecting services relevant for the user (Goker and Myrhaug, 2002).

From the above meaning it follows that a context-aware system or process is one that considers also situation data for achieving a desired objective. In a well-known definition, Dey (2001) states that the context is "*any information that can be used to characterize the situation of an entity*". In the case of movement data analysis the entity is the moving object while the context data are any data that can characterise the situation in which the movement takes place. They can describe the characteristics of the moving entity such as the gender of the moving human or the type of the moving vehicle. Movement context data can describe the time at which the movement takes place such as whether it is night or daytime. Movement context data can also describe the geographic environment of movement such as the type of road on which a vehicle moves and other entities moving in close proximity.

### 2.4.2 Semantic enrichment

The process of adding at some step of movement data analysis information that can support interpreting movement patterns is known as semantic enrichment (Baglioni et al., 2009). The semantic enrichment can be done by deriving the information directly from the movement data or by incorporating context data. Semantic enrichment directly from movement data has been done, for example, by detecting stops along the movement using time and distance thresholds (Li et al., 2008). In another example, Buchin et al. (2013) identified movement states (e.g., flight, stopover) directly from movement data based on movement parameters (e.g., heading angular range, minimum speed).

Different approaches have been followed for semantic enrichment using context data. For example, Baglioni et al. (2009) encode Points of Interest (POIs) as classes in an ontology and then run a reasoner on the ontology to infer the interpretation of movement patterns. Yan et al. (2010) use spatial joins for integrating trajectory episodes with regions of interest (e.g., land uses). They also integrate trajectory episodes with lines of interest (e.g., roads), and points of interest (e.g., hotels) using map matching, and hidden Markov model respectively. The system developed by Yan et al. (2010) provides a good general framework for semantic enrichment. The framework is general in the sense that it supports the integration of geographic context elements of different dimensions (area, polyline, and point). However, the framework does not consider any dynamics of the context. For example, it can relate a trajectory episode with a road but without considering any change on the road (e.g., its traffic congestion level). Furthermore, the framework can relate only trajectory episodes with context objects that spatially overlap with them. In some cases a context object influences the movement of an entity which is not in its close proximity. For example, a taxi moving towards a road junction may learn from a distance that the junction is congested and decide to bypass it. Even if the taxi and the junction are not in close proximity, in such case the trajectory of the taxi should be related to the junction to analyse the movement pattern of the taxi.

### 2.4.3 Context-aware analysis of movement data

Since very little can be achieved in semantic enrichment of movement data and patterns directly from movement data alone, the trend is a context-aware analysis of movement data; that is, an analysis integrating the movement data with context data. In this line, Wachowicz et al. (2013) integrated context data characterising visitors of a recreational area to support interpretation of their flock patterns. Context data were in the form of contextual variable such as the age category of the visitor, whether he is a local, and the frequency of visit.

The larger literature on context-aware analysis of movement data is about the integration of context data describing the geographic environment of the movement. Considering that there are two objects being related, the moving entity and the object concerned by the context data, this context integration has been commonly studied under the concept of *movement interaction*. The movement is conceptualised based on the interactions that happen between the moving entity and the environment in which it moves (Orellana and Renso, 2010). For example, Orellana et al. (2012) integrated pedestrians' movement data with point locations of attractions and facilities in a park to explore the interactions of pedestrians with a recreational area. This context data integration supported an interpretation of the patterns of stopping and sequences of stopping of pedestrians. Interactions between a moving entity and its environment have been widely studied in ecology.

#### Movement data analysis in movement ecology

The field of ecology deals with the relations of leaving organisms to one another and to their physical surroundings. This relation is reflected in the movement of the organisms. Movement ecology seeks to understand fundamental questions about movement of living organisms in the context of their interactions among themselves and with the geographic space.

The study of interactions between a moving animal and the geographic space enable understanding of the influence of the environment on the movement pattern of the animal. For example, Safi et al. (2013) integrated the movement data of migrating birds with weather data to assess wind effect on the birds' flight pattern determined by the flight direction and speed. In another study relating the movement data of roe deer with climatic (snow) and topographic (slope) data (Cagnacci et al., 2011), the authors explored the impact of these factors on the patterns of migration of the roe deer. They observed that the migration occurrence depended on winter severity and topographic variability.

From the study of interactions between a moving animal and the environment, it is also possible to derive information about the geographic space. In this line, real time traces of tagged animals enable documenting ongoing environmental changes. This is done, for example, by analysing the change in how species with large movement use their habitat (Kays et al., 2015). In another example, instrumented seals have been used to monitor changes in the Southern Ocean while common methods based on Argo profilers would face the obstacle of sea ice (Roquet et al., 2013). Conductivity-temperature-depth satellite relay data loggers, originally used to improve understanding of seal foraging strategies, provided as a by-product a cost-effective method of sampling hydrographic properties in Southern Ocean regions.

The study of interactions between a moving animal and one or more other animals supports understanding of the effect of the presence of one animal on the movement pattern of the other, and the relationships between the animals. For example, Pettit et al. (2013) studied the movement of a group of homing pigeons and found that momentary changes in velocity of one pigeon were a response to the neighbour's orientation and position. This enabled further to understand how a route decision emerges from the interaction between the group members. Different animal behaviours (e.g., familial bonds and mating behaviour) or relationships between animals (e.g., predator-prey) have been interpreted from their interactions. The analysis of movement data of wolves and moose (Eriksen et al., 2009) led to identifying a movement pattern of prey seeking and predator avoidance. The study of interactions between three maned wolves (an adult pair and their juvenile female offspring) revealed a movement pattern showing constant close proximity (Bandeira de Melo et al., 2007). This movement pattern was interpreted as a strategy to maintain familial bonds. In another study (Long and Nelson, 2013), the analysis of the movement of a male-female bear combination revealed a highly cohesive movement pattern at specific periods, which was interpreted as mating behaviour.

The interaction between animals has been widely studied in ecology. Long et al. (2014) distinguish static interaction from dynamic interaction. Static interaction is defined as the joint space use between two individuals as indicated by an index of home range overlap. Dynamic interaction indicates the relation between the movements of two individuals. While in the static interaction the time of space use is not considered, the dynamic interaction implies a simultaneous space use. Dynamic interactions have received a high research interest leading to different methods for measuring them. As shown by the evaluation of these methods (Long et al., 2014; Long, 2015), they are mainly defined based on the proximity (spatial and/or temporal) between the recorded locations of the individuals.

In the study of animal movement, context data are commonly represented as a field covering the whole study area as obtained through remote sensing or point interpolation techniques. They represent environmental conditions that can influence the movement of the animal. Discovering and understanding this influence is the main objective of the study. With increasing access to large repositories of such environmental data in different formats, there has been important work in movement ecology focusing on linking animal tracks to these data. For example, the *Environmental-Data Automated Track Annotation (Env-DATA)* system (Dodge et al., 2013) allows integrating animal movement data with environmental data (e.g., weather, topography ...). Different formats such as GRIB, GeoTIFF, and ASCII from different satellite sources are supported. This system has been used to examine the influence of the wind on the flight of Galapagos Albatrosses (Dodge et al., 2013). Likewise, Dodge et al. (2014) integrated Turkey vulture tracks with NDVI[1] data from MODIS[2] to examine the correlation of Albatrosses' movement pattern and the availability of food for their feeding.

### 2.4.4 Sources of geographic context data for movement analysis

Data about the geographic context in which the movement takes place can be obtained in many different forms and from different sources. The important requirement is that they are geo-referenced so that they can be overlaid with the movement data. Furthermore, in case of dynamic context the time period in which the context data

[1]Normalized Difference Vegetation Index

[2]MODerate resolution Imaging Spectroradiometer (https://modis.gsfc.nasa.gov/data/)

was collected should coincide with that of the movement data. The main sources of geographic context data are discussed next:

**Other georeferenced datasets of the same area**: These data can be obtained from mapping agencies (e.g., Ordnance Survey[3] , USGS[4] ...) and public data repositories on the Internet in different forms. For example, vector data such as Points of Interest (POIs), infrastructure (e.g., road network) and administrative subdivisions can be retrieved from the OpenStreetMap[5] and GADM[6] websites. Georeferenced data on environmental conditions, especially for the analysis of animal movement, can be obtained from MoveBank[7] . Likewise, weather data can be obtained from weather services (e.g., European Centre for Medium-Range Weather Forecasts (ECMWF)[8]).

**Geo-social data**: Geo-social media platforms enable users to share georeferenced data about locations where they are or have been. These data include geographic coordinates and descriptive information, which makes them rich in semantics. A process called geo-tagging enables associating media data such as photographs and images with geographical identification (Spinsanti et al., 2013). Geo-social data are available in many different forms. Check-ins from Foursquare[9] embed descriptions of places which can inform about activities carried out in the places for example. Photos and associated description from Flickr[10] can inform about events occurring at the locations where they have been taken. Geo-tagged tweets, which are short texts posted on Twitter[11] inform about the situation at the locations they are associated with. For example, a tweet may inform about the current situation of a rugby match that the sender of the tweet is watching. Tweets may be associated with other type of data (e.g., pictures) which make them richer in semantics.

Geo-social media platforms provide APIs (Application Programming Interfaces) that are used to access the data. However, most of these data are associated with hard access restrictions especially regarding the amount of data that can be continuously acquired and the access to historical data. For example, Foursquare sets a limit of 5000 requests per hour and does not allow access to data of a selected user by another

[3]https://www.ordnancesurvey.co.uk/
[4]https://www.usgs.gov/
[5]https://www.openstreetmap.org/
[6]http://www.gadm.org
[7]https://www.movebank.org/
[8]http://www.ecmwf.int/
[9]https://foursquare.com/
[10]https://www.flickr.com/
[11]https://twitter.com/

user. Likewise, Twitter allows at maximum 150 unauthenticated and 350 authenticated requests per hour. Also, geo-social data require an appropriate method (e.g., natural language processing methods (Dashdorj et al., 2016)) to automate the extraction of the semantics they contain.

**Data from other sensors**. Geographic context data are sometimes acquired using other sensors that sense the environment where the object is moving. The sensors can be in the same device that records the movement or in a different one. For example, a smartphone embeds multiple sensors, some of which can provide geographic context data for analysing the movement of the individual carrying the smartphone. The Bluetooth sensor records the presence of other Bluetooth-enabled devices and hence presence and number of other static or moving objects. The temperature sensor (or thermometer) records the ambient temperature in the location where the phone carrier moves while the microphone can record the noise level.

There are smartphone-based applications that access the GPS and other sensors such that they can collect movement data and context data at the same time. For example, “CenceMe” (Miluzzo et al., 2007) can record the user’s movement and the temperature and noise in his surrounding using smartphone-based sensors. Likewise, “LifeMap” (Chon and Cha, 2011) can use GPS, thermometer, and Bluetooth sensors embedded in a smartphone to record the position of a moving user and attach the ambient temperature and the number of Bluetooth-enabled devices in his surroundings. Additional sensors can also be deployed while recording the movement of animals. For example, Bleisch et al. (2014) present an analysis of fish movement data and associated context data. The fish movement was recorded along with river environmental data including water temperature and water level.

**Trajectories of other moving objects**. For each moving object, trajectories of other objects that move in its spatial and temporal proximity constitute geographic context data because the other objects can influence its movement (e.g., by slowing it down). This is the case of group movement patterns presented in section 2.3.2. Therefore, while analysing the movement of a vehicle which is part of a convoy or the movement of a bird which is part of a flock of birds, the trajectories of other convoy members respectively flock members are considered as geographic context data. The trajectories of other moving objects need not be in the same dataset. For example, trajectories of buses operating in the same area and same period as taxis constitute geographic context data for the analysis of the trajectories of these taxis.

**Other sources**. Information about events at specific POIs can be obtained from event listing, reports and news webpages. This information forms semi-structured or unstructured data that can be parsed and used as geographic context data for the places and locations about which they are published. Analyst' background knowledge about places and locations can also be used as geographic context data linked to movement data in the same locations and places. Like most of context data from event listing, reports and news, context data from analyst's background knowledge is not structured. The unstructured nature of these types of context data makes it hard to integrate them into movement data analysis.

## 2.5 Summary

This chapter presented the basic concepts and state-of-the-art pertaining to the themes dealt with in the following chapters. It provides an overview of movement data and movement data analysis. Since movement data are a special type of geographic data, the chapter introduced geographic data and relations between geographic objects. Special focus of the discussion has been put on qualitative relations including spatial (e.g., topological) and spatio-temporal relations (e.g., QTC relations). The relations discussed are defined for specific cases (e.g., a pair of static objects, a pair of moving objects). Therefore, these relations are used in the following chapters as building blocks for defining new relations appropriate to describe a moving object and the geographic context in general. The notion of conceptual neighbourhood graphs (CNG) used to structure the relations as discussed in this chapter is used in the following chapters to structure new relations that are introduced. The qualitative reasoning introduced in this chapter is proposed in the following chapters as one analysis approach possible on the new relations introduced. The chapter discussed the nature of movement data including their different forms and inherent issues associated with them. These issues are taken into account in the following chapters while processing the evaluation data. Though the methods presented in the next chapters are general, they are evaluated on a specific form of movement data, namely GPS-based data, and at a specific scale.

This chapter presented the state-of-the-art on movement data analysis. It introduced the KDD process with emphasis on the data mining and pattern interpretation/evaluation phases. The following chapters build on the analysis approaches discussed in this chapter to propose an analysis framework taking into account a dynamic geographic context. In particular, the concept of movement interactions, presented in this chapter, has been generally defined between moving entities. This concept is generalised on different types of geographic context elements defined in the following chapters. Some of the data mining methods presented in this chapter (e.g., clustering and classification) are used to implement the analysis framework proposed in the following chapters. Since the focus of this thesis is on a context-aware analysis of movement data, this chapter discussed existing work in this area. The static context, generally dealt with in the work presented in this chapter, is taken as reference in the following chapters for identifying the factors that make a geographic context a dynamic one. In this chapter, I surveyed the sources of context data that can support movement data analysis. This survey contributes to answering the research question on the approach of integrating context data into movement data analysis. The visual analysis methods discussed in this chapter complete the discussion of the spatio-temporal nature of movement data. However, a detailed discussion of these methods is beyond the scope of this thesis. Therefore, computational methods and simple visualisations are adopted in the following chapters.

# Chapter 3

# A conceptual model of movement context and interactions

In this chapter, different concepts for modelling the context and relating movement to it are presented. The chapter starts with a discussion of different categories of contexts for moving objects, and then presents different approaches for modelling the context. After, the discussion of the context in general, the chapter focuses on a dynamic geographic context and presents an approach for relating it to movement.

Some authors explained the concept of context by establishing categories of context. For instance, Goker and Myrhaug (2002) proposed a generic user context comprising five sub-categories: *environment* context, *personal* context, *task* context, *social* context, and *spatio-temporal* context. The environment context is made of entities that surround the user. It includes for example things, services, temperature, light, humidity, noise, and persons. The personal context category comprises internal elements that characterise the person, which can be for instance the blood pressure, weight, and expertise. The task context category describes what the persons are doing in this user context. The description can include, for example, the goals and activities. The Social context describes the social aspects of the current user context with information such as friends and neighbours. The spatio-temporal context describes aspects relating to the time and the spatial extent for the user context. The spatio-temporal context can contain attributes such as the time and the spatial extent of the environment. It is important to mention that it is hard to draw a clear boundary between the environment and the spatio-temporal categories. In a recent study, Buchin et al. (2014) defined the *geographic* context and categorised it into network, land cover, obstacles, terrain, and ambient attributes.

The focus of this thesis in regards to context is on the environment of a moving entity. This corresponds to some elements of the "environment" and the "spatio-temporal" sub-categories of the user context proposed by Goker and Myrhaug (2002). These are the elements that have locations relative to the surface of the Earth. They are found in the different categories of the geographic context proposed by Buchin et al. (2014). Similar to the context studied by Buchin et al. (2014), the context made of these elements is also called geographic context in this thesis. The definition of the context as considered in this thesis is given in the following section.

## 3.1 Modelling the geographic context

There are several objects that have a geographic location at some time and hence are part of the geographic context. These elements of the geographic context can be put into the following categories: *space, static objects, moving objects*, and *events.*

**Space**. The geographic locations where the moving object passes have some properties which characterise them. For instance, the movement space may be a road network where different locations may be on different types of roads (e.g., highway and street) and different road segments may be on different land use types. The movement space may be a park where some locations are on open areas while others are on dense forest areas. Environmental conditions such as wind speed and temperature at different locations are other examples of geographic context of category "space".

**Static objects**. Some objects have a fixed geographic location (i.e., that does not change with time) and the movement may be related to them in some way (e.g., the moving object stops at or passes by them). Examples of static objects are bus stops, traffic lights, zebra crossings, and information boards. Static objects have been widely considered in movement analysis under the general term of Points of Interest (POIs).

**Moving objects**. Unlike static objects, moving objects change their location with time. Objects in this category are, for example, other moving cars, animals, or pedestrians crossing the road followed by a car.

**Events**. In a general sense, an event is defined as "a temporally bounded happening" (Galton, 2012). Under this general definition, the term event has been used to mean an episode of movement with a specific characteristic (e.g., passenger pickup event of a taxi (Ding et al., 2015) and slowing down event (Vandecasteele et al., 2014)). It can be used also to refer to happenings without explicit physical location (e.g., online student registration). As this thesis is concerned with a geographic context, events considered are temporally bounded happenings with physical geographic locations. This type of events has particularly attracted attention for automatically detecting them based on people's response that they trigger. For example, Polous et al. (2015) developed a general framework for detecting these events from social media. The scope of this thesis is limited on events that can limit or motivate objects to move. For example, social events (e.g., a concert, a football match ... ) that attract a considerable number of people and non-social events (e.g., traffic congestion, roadwork) that limit movement are in the scope of this thesis. On the other hand, events such as a plane crash and a volcanic eruption that occur in a remote place are not in the scope of this thesis as they generally do affect the movement of objects.

#### Properties of geographic context elements

Geographic context elements have attributes, which provide the semantics needed to support movement analysis. For example, a road segment has a type attribute, which indicates the category of road being used for movement. In addition to thematic attributes attached to a geographic context element, this thesis considers three attributes: *location* and *extent* derived from the spatial characteristics of the element, and *lifespan* derived from its temporal characteristics. The *location* attribute indicates whether the geographic location of the context element is `fixed` or `changing` while the *extent* attribute indicates whether its size and/or shape are `fixed` or `changing`. The *lifespan* attribute indicates whether the existence of the context element is `bounded` or `extended` beyond the period being studied. In other words, the value `bounded` for the lifespan attribute means that the element exists for a time interval shorter than the time covered by the study such that the bounds of the time interval are known or important in the study. The value `extended` means the opposite.

Based on the attributes of its elements, a geographic context may be characterised as *static* or *dynamic* (see Table 3.1). A geographic context is dynamic if at least one of the following four conditions is fulfilled: 1) the *location* attribute of the context element has a value `changing`, 2) the *lifespan* attribute has a value `bounded`, 3) the context element has a time-varying thematic attribute (also called a dynamic attribute) which is being studied, 4) the *extent* attribute has a value `changing`. A geographic context which is not dynamic is static. In other words, if a geographic context changes with time it is dynamic, otherwise static. It follows that a moving vehicle (a context element of type *moving object*) makes always a dynamic context because the location of the vehicle changes with time. Likewise, a football match (a context element of type *event*) makes always a dynamic context because it has a bounded lifespan; i.e., it exists only for a short time compared to the period covered by the analysis. On the other hand, a road segment (as a context element of type *space*) makes a static context if the interest is only on its location and the road type. The same road segment makes a dynamic context in case the interest is on the level of congestion on it. Likewise, a school is a static context element if the interest is only on its location, but it is a dynamic context if the interest is on the number of people leaving it.

| **Context element type** | **Attributes** | | | | **Context characteristics** |
|---|---|---|---|---|---|
| | **location** | **lifespan** | **extent** | **thematic attributes** | |
| static object | fixed | extended | fixed / changing | not varying / varying | Static / dynamic |
| moving object | changing | extended | fixed / changing | not varying / varying | dynamic |
| event | fixed / changing | bounded | fixed / changing | not varying / varying | dynamic |
| space | fixed | extended | fixed / changing | not varying / varying | Static / dynamic |

Table 3.1 Properties of a geographic context

The integration of the geographic context into computational movement data analysis requires that the context be represented in some form supported by existing computational systems. Since a geographic context is a geographic phenomenon, standard geographic data models (i.e., object-based and field-based) are selected in this thesis. A combination of different geographic data models and types of context elements results in different possible models of the geographic context as shown in Figure 3.1. Figure 3.1 shows each type of geographic context element in different models with an example of a context element – moving object pair. The moving object is represented by its trajectory (see the dotted line).

The choice of the model to use for the geographic context depends on the application case and the application field. For example, in ecology the study area is generally wide, remotely located, and the movement is less restricted (e.g., seabirds commuting between nesting and foraging areas (Tarroux et al., 2016)). In such situation remote sensing techniques appear as the cost-effective method to collect geographic context data with full spatial coverage. Since this method produces data in raster format, which fits well the field-based model, this model is especially adopted in ecology. Hence, in the study of the influence of the wind conditions on the flight behaviour of an Antarctic seabird (Tarroux et al., 2016) the field-based model was used. The wind data, as geographic context data, were represented as a field covering the wide area linking the nesting and foraging areas. Unlike in ecology, the movement of people in transportation is generally restricted to road networks which are commonly represented using object-based models. The study area in transportation is also relatively small and application cases require a higher spatial resolution than in ecology (e.g., for relating transportation vehicles to road segments and road infrastructure elements (Wang et al., 2014)). Therefore, in transportation the object-based models are commonly adopted. Furthermore, for a given context element different aspects of multiple representation (Timpf and Devogele, 1997) also contribute to the choice of the model to use. Among others, the scale considered by the application determines the appropriate model to use.

*Static objects* are modelled as shown in the first column of Figure 3.1. A static object can be for example a school considered with respect to movement of a car on the road network of the whole city, in which case it is better modelled as a point. The same movement of a car can be related to a river going through the city. In this case, the river is a static object better modelled as a linear object. For a bus operating only in the neighbourhood of a university campus, the university campus is a static context element better modelled as a polygon. It can be further represented in a field-based model; for example if the data are collected using remote sensing techniques.

*Moving objects* can be modelled in different ways as shown in the second column of Figure 3.1. A pedestrian is a moving object because his position changes with time. With respect to movement of a car, a pedestrian can be modelled as point at any time instant of his movement. On the other hand, a sea wave appears as a moving one dimensional object. Therefore, with respect to ship movement it can be better represented as a line. An oil spill on the sea appears as a moving surface. Therefore, with respect to ship movement it can be better modelled as a polygon and possibly as a field.

*Events* may have different extents and hence modelled differently as shown in the third column of Figure 3.1. For instance, with respect to the movement of a car on the city-wide road network, a concert in a specific theatre can be modelled as a point with same location as the theatre. A demonstration taking place along a street has a linear extent at any time instant during its progress. Therefore, with respect to the car movement a street demonstration can be better modelled as a line. Some events

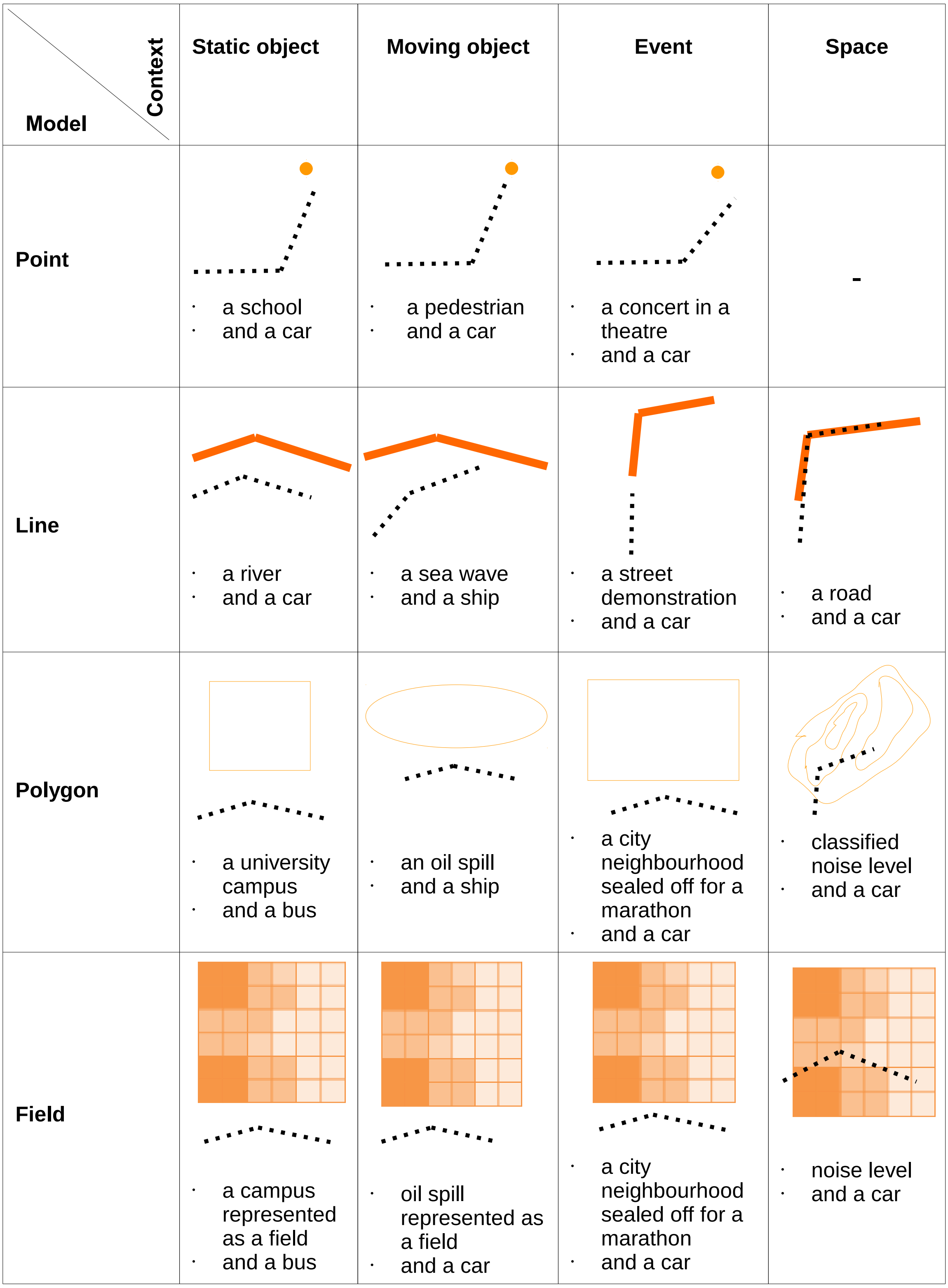



Fig. 3.1 Different types of geographic context, their different models and examples

can take place on a wide area or around a wide area such that the whole area may be considered as the event location. For example a city marathon may be organised to use streets in and around a neighbourhood such that the whole neighbourhood is sealed off for the duration of the event. In this case, with respect to a car movement in the city, the marathon can be better modelled as a polygon, and possibly a field, with the extent equal to the neighbourhood it covers.

*The Space* in which movement takes place is non-zero dimensional because it includes at least two different positions of the moving object. Therefore, as shown in the fourth column of Figure 3.1 it cannot be modelled as a point because a point is zero-dimensional. On the scale of a city, the road on which a car moves is better modelled as a line. The space can also be extended on a surface. Environmental conditions attached to the geographic extents in which they are measured form also the space. For example, the level of noise in different urban areas where a car moves forms a geographic context. The noise level can be modelled as a field or sometimes as polygons representing different noise classes.

In this section, the geographic context in general has been described and a general approach for modelling it presented. The next section is about a dynamic geographic context because this is the focus of the thesis. However, the terms "geographic context" and "context" will continue to be used where it is not important to differentiate between a geographic context which is dynamic and one which is static. The different properties that make a geographic context dynamic may imply specific considerations in some analysis cases. For example, the change of the extent of a context element implies that the uncertainty should be considered while referring to its boundary. Nevertheless, the next sections present a general approach applicable irrespective of the type of changes that the context element undergoes. The particularities due to types of changes can be handled as slight adjustments while implementing the general approaches.

## 3.2 Relating the movement to a dynamic geographic context

In this section I describe an approach for relating movement to a dynamic geographic context. I consider the movement to be represented by its trajectory made of a chronologically ordered list of points recorded along the movement. The geographic context is modelled according to one of the models presented in section 3.1. The approach proposed for relating movement to the geographic context is based on the concept of *movement interactions*. I first introduce this concept and then discuss how it is applied.

### 3.2.1 Movement interactions

Both the moving object and the geographic context are geographic phenomena. It implies that at some point in time there exists a spatial relation between them (e.g., a car is outside of a park). As the object moves, this spatial relation may change (e.g., the car that was outside of the park is now inside). We call *movement interaction* (or simply *interaction*) the process in which a spatial relation between a moving object and the geographic context changes over time. This concept of interactions, which has also been used in (Orellana and Renso, 2010), has been called *spatio-temporal relations* by Andrienko, Andrienko and Heurich (2011) and *patterns* by Dodge et al. (2008).

The concept of "interaction" generally means a reciprocal action or influence. This makes it suitable for use in addressing the two-fold objectives of the thesis: to understand movement from the context and to learn about the context from movement. The effect of the context on the moving object is expected to be observed as a different movement pattern when the context or its properties change. For example, an animal initially running in zigzag to escape from a predator will stop or move normally when the predator goes away. The effect of the moving object on the context is reflected in the characteristics of the context element which are expected to be different in case there is a moving object compared to the case there is none. For example, walkers form trails on a deformable terrain which can in turn influence the motion of walkers on this terrain (Helbing et al., 2001).

Movement interactions can be defined based on the change of any of the three basic spatial relations discussed in section 2.1.2 (namely, topological, directional, and distance relations). For example, we can define an interaction "bypassing" based on a sequence of changes of distance relations and changes of directional relations. In this case, the moving object decreases its distance to the context element till a certain minimum value, then changes its direction while in the proximity of the context element, and finally increases the distance from the context element. The approach I follow to define interactions is closely related to some previous work on modelling the motion of objects. The Qualitative Trajectory Calculus (Van de Weghe, 2004), QTC in short, defines relations between two moving point objects. The basic version (QTC Basic or $QTC_B$) defines the relations by comparing differences in distance between the objects over time. Another closely related work is (Salamat and Zahzah, 2012) which defines spatiotemporal relations comparable to the movement interactions defined in this thesis. Different spatiotemporal relations are defined based on a combination of topological and directional relations that hold over a time interval.

In this thesis I define movement interactions based on the change of distance relations only in a way closely similar to the approach of $QTC_B$ in (Van de Weghe, 2004). QTC relations (Van de Weghe, 2004) and the spatiotemporal relations presented in (Salamat and Zahzah, 2012) are defined between two moving point and two-dimensional objects respectively. Unlike these relations, the movement interactions defined in this thesis concern one moving point object and one context object which may be moving or not. Also, I define the interactions in a systematic way based on a continuous motion such as in (Van de Weghe, 2004) and a categorisation of objects involved. This allows taking into account specific properties of involved objects. The focus is on a small set of basic interactions and how they vary among different context element types. In order to define the basic interactions I assume the change of distance between the moving object and the context element to be continuous and linear; that is, the distance decreases progressively to become zero before increasing progressively. In other words, the distance can be represented by a signed number from the set of real numbers ($\mathbb{R}$) where negative (-) means "decreasing" while positive (+) means "increasing". The distance decreases till the minimum value (theoretically zero) before increasing. I define basic interactions for three cases based on the context element involved. The first case concerns interactions with a zero-dimensional (0D) context element having a fixed location. The second case concerns interactions with a one-dimensional (1D) or two-dimensional (2D) context element having a fixed location. The third case concerns interactions with a moving zero-dimensional geographic context element.

The movement interactions are defined and named based on the semantics of motion verbs such as those presented in (Asher and Sablayrolles, 1994). The notations that are used to formalise the definition of interactions are introduced in the following.

- `A` represents a moving object.
- `C` represents a context element with which A interacts.
- `d(A,C,t)` denotes the distance between `A` and `C` at time instant `t` sampled for position recording.
- `nParam` denotes a nearness parameter. That is; when `d(A, C, t) <= nParam`, `A` is said to be at `C`. For a given object, this parameter determines a small area around it in which all positions can be approximated by the position of the object. It serves as a tolerance area for locating another object which can be

considered to be in contact with the target object or in a negligible distance from it. The parameter can be set based on domain knowledge. For example, on a road network it can be set to the distance between the object and the nearest road junction, or simply the positioning accuracy (e.g., 20 m for GPS positions).

- $\mathtt{S_{min}}$ denotes the minimum amount of time that a valid stop of the moving object should last.
- $\mathtt{td(A,t_1,t_2)}$ denotes the distance travelled by $\mathtt{A}$ during the time interval $\mathtt{[t_1,t_2]}$.

**Case 1: A moving object and a zero-dimensional geographic context with a fixed location**

In the case of a moving object and a zero-dimensional geographic context element with a fixed location, five basic movement interactions are defined: *approaching, arriving, stopping, leaving*, and *moving-away*. These interactions are depicted in Figure 3.2. The moving object is drawn using three small circles with different radii to indicate the progress of movement. The bigger the circle is the more recent is the position it represents. So, the biggest circle represents the latest position while the smallest dotted circle represents the earliest position. The five movement interactions depicted in Figure 3.2 are defined next.

**Definition 1 (stopping)** . *A stopping interaction happens when the initially moving object stays in the neighbourhood of the context element for a certain time threshold.* This is formalised as follows:

$$\exists t_{\mathrm{i}}, t_{\mathrm{j}}, t_{\mathrm{k}} | t_{\mathrm{i}} < t_{\mathrm{j}} < t_{\mathrm{k}}$$
$$\forall t, t_{\mathrm{j}} < t < t_{\mathrm{k}} : d(A, C, t) \leq nParam \wedge t_{\mathrm{k}} - t_{\mathrm{j}} \geq S_{\mathrm{min}}$$

**Definition 2 (approaching)** . *The approaching interaction is observed when the distance between the moving object and the context element decreases.* The approaching interaction is formalised as follows:

$$\exists t_{\mathrm{g}}, t_{\mathrm{h}} | t_{\mathrm{h}} > t_{\mathrm{g}}$$
$$d(A, C, t_{\mathrm{h}}) - d(A, C, t_{\mathrm{g}}) < 0 \wedge d(A, C, t_{\mathrm{h}}) > nParam$$

**Definition 3 (moving-away)** . *The moving-away interaction is observed when the distance between the moving object and the context element increases.* This is formalised as follows:

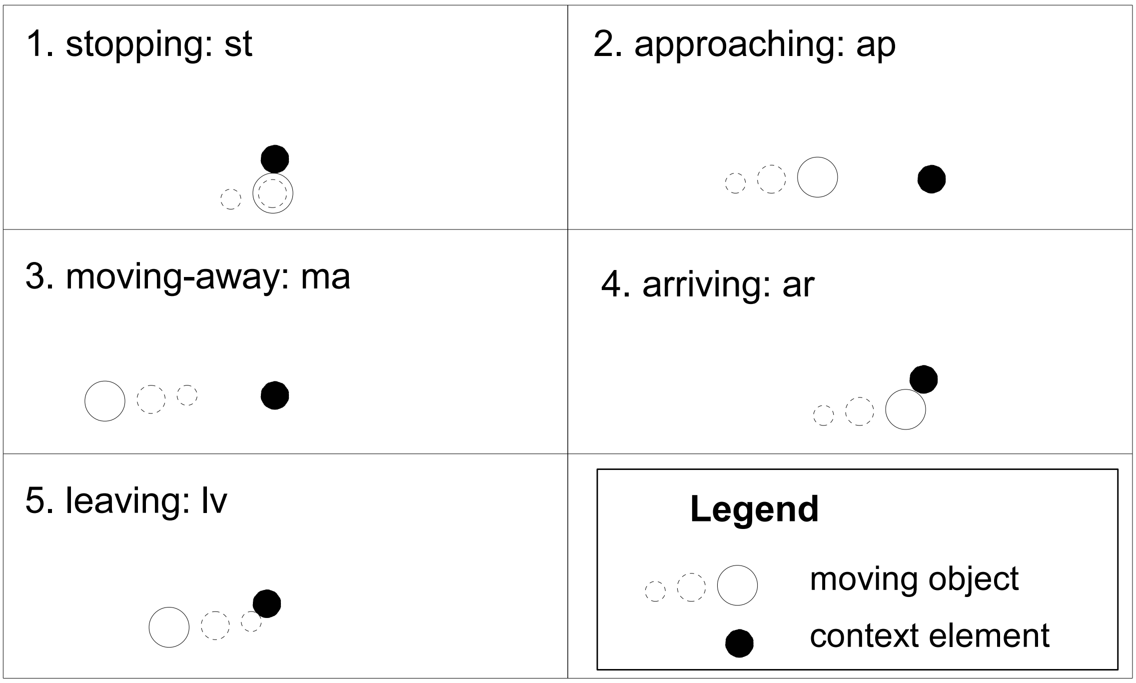



Fig. 3.2 Basic interactions between a moving object and a 0D context with a fixed location

$$\exists t_{\mathrm{l}}, t_{\mathrm{m}} | t_{\mathrm{m}} > t_{\mathrm{l}}$$

$$d(A, C, t_{\mathrm{m}}) - d(A, C, t_{\mathrm{l}}) > 0 \wedge d(A, C, t_{\mathrm{l}}) > nParam$$

**Definition 4 (arriving)** . *Arriving is observed when the distance between the moving object and the context element deceases till the object reaches the neighbourhood of the context element.* This nteraction is formalised as follows:

$$\exists t_{\mathrm{i}}, t_{\mathrm{j}} | t_{\mathrm{i}} < t_{\mathrm{j}} \wedge j = i + 1$$

$$d(A, C, t_{\mathrm{i}}) > nParam \wedge d(A, C, t_{\mathrm{j}}) \leq nParam$$

**Definition 5 (leaving)** . *Leaving is observed when the distance between the moving object and the context element increases such that the object initially located in the neighbourhood of the context element is no longer in its neighbourhood.* This interaction is formalised as follows:

$$\exists t_{\mathrm{i}}, t_{\mathrm{j}} | t_{\mathrm{i}} < t_{\mathrm{j}} \wedge j = i + 1$$

$$d(A, C, t_{\mathrm{i}}) \leq nParam \wedge d(A, C, t_{\mathrm{j}}) > nParam$$

It is important to note that the set of basic interactions does not constitute an exhaustive list of interactions. Other interactions can be defined by considering the variation of other spatial relations or even by combining some basic interactions. For example, an interaction "passing" can be defined as sequence of arriving and leaving. Interactions such this can be considered as second degree interactions because they can be defined as a composition of basic interactions.

#### Case 2: A moving object and a 1D or 2D geographic context with a fixed location

A context element of type "space" extends over an infinite set of point locations which make either a linear feature (1D) or an areal feature (2D). Some of the basic interactions defined for the case of a 0D context with a fixed location are still valid for the current case. The names of these interactions can change for better reflecting the higher dimensional nature of the context. For example, the *arriving* interaction can be called "entering" while the *leaving* interaction can be called "exiting". Alternative terms for naming interactions can be identified based on the semantics of motion verbs in natural language such as presented in (Asher and Sablayrolles, 1994). In this thesis, the same naming is kept in the different cases as long as the semantics of the interaction is the same. The higher dimensional nature of the context element in this case also introduces a new basic interaction, which is named *passing*. It is important to note that *passing* on a 1D or 2D context element is not equivalent to a sequence of *arriving* and *leaving* because in this case there is a time interval between the two.

**Definition 6 (passing)** . *The passing interaction occurs when the distance between the moving object and the context element of type space changes as follows: it is initially greater than a specified nearness parameter, then it drops and remains below a certain threshold close to zero within some time interval before increasing again above the value of the nearness parameter.*This is formalised as follows:

$$\exists t_{\mathrm{k}}, t_{\mathrm{l}}, t_{\mathrm{x}}, t_{\mathrm{y}} | t_{\mathrm{x}} < t_{\mathrm{k}} < t_{\mathrm{l}} < t_{\mathrm{y}}$$

$$d(A, C, t_{\mathrm{x}}) > nParam \wedge d(A, C, t_{\mathrm{y}}) > nParam \wedge$$

$$\forall t, t_{\mathrm{k}} \leq t < t_{\mathrm{l}} : d(A, C, t) < nParam$$

The basic interactions in case of a 1D or 2D context with a fixed location are depicted in Figure 3.3. In Figure 3.3, the context element is represented by a rectangle to reflect its higher dimension compared to the zero-dimension of the previous case. However, the actual context element can be a linear feature (e.g., a road for a car or a

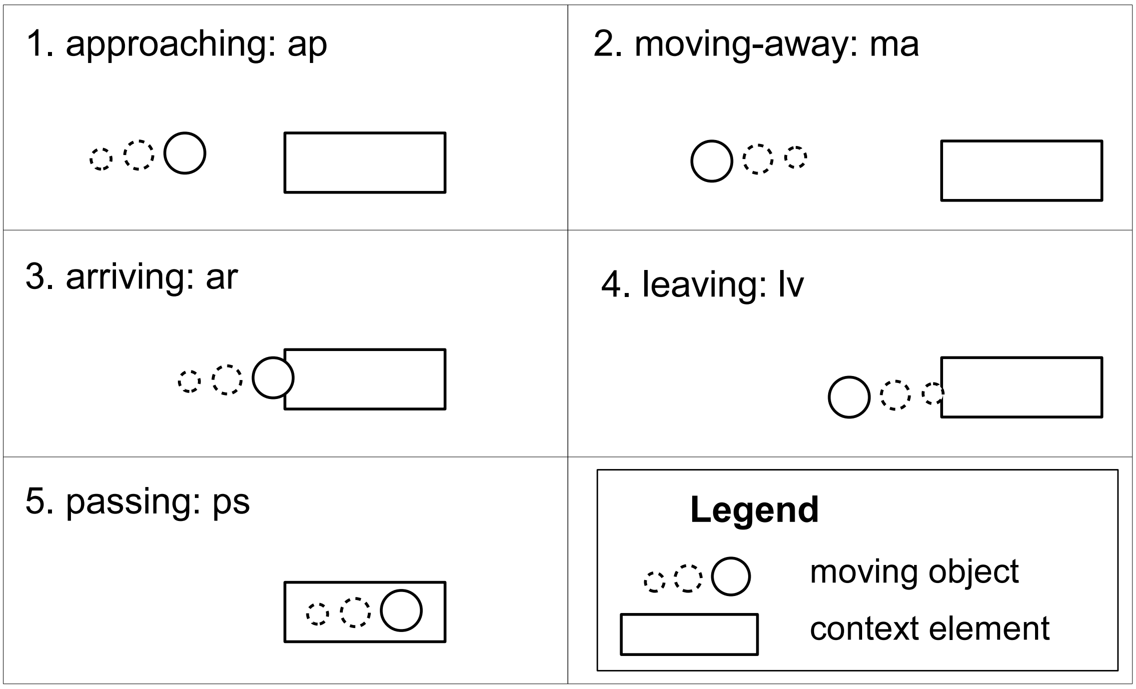



Fig. 3.3 Basic interactions between a moving object and a 1D or 2D context with a fixed location

canal for a boat) or an areal feature (e.g., city neighbourhood). The symbol used to represent the moving object has the same meaning as in the first case.

**Case 3: A moving object and a moving zero-dimensional geographic context**

In case the context element also moves, interactions basically similar to those defined in previous cases may have a different semantics. For example, the term *arriving* is appropriate in case there is a fixed location at which the moving object arrives. In the current case a new term, which reflects the movement of both objects, is used. The new term, named *encounter*, means that the moving object reaches the context element (which is also moving). The opposite interaction, which is similar to *leaving*, is called *separating*. These two interactions represent both the cases of same and different directions. If the objects encounter and stay together in the same location for some time, the interaction is not called *stopping* but *meeting*. The *meeting* interaction reflects the fact that the objects arrive in a location at the same time. Since both the target and context objects move, other interactions can differentiate whether they move together or separately. This leads to two new interactions: *jointly-moving* and *separately-moving*.

The basic interactions considered for the case of a moving context element are illustrated in Figure 3.4. In the figure, the symbols used have the same meaning as the symbol for the moving object in the preceding cases. Though the meaning of the two symbols gives an impression of a specific movement direction these interactions represent both cases of same and different movement directions. A specific direction is selected simply for making the illustration easy to understand. The new interactions are defined next.

**Definition 7 (meeting)** . *The meeting interaction occurs when the target moving object enters the neighbourhood of the moving context element and they stay together for a specified time threshold. In other words, the distance between the target moving object and the moving context element deceases below a predefined nearness parameter. The distance remains below the nearness parameter for some time. The distance travelled while staying together does not exceed the value of the nearness parameter, which means that once they are together the two objects do not move.* This is formalised as follows:

$$\exists t_{\mathrm{i}}, t_{\mathrm{j}}, t_{\mathrm{k}} | t_{\mathrm{i}} < t_{\mathrm{j}} < t_{\mathrm{k}}$$

$$d(A, C, t_{\mathrm{i}}) > nParam \wedge$$

$$\forall t, t_{\mathrm{j}} \leq t \leq t_{\mathrm{k}} : d(A, C, t) \leq nParam \wedge t_{\mathrm{k}} - t_{\mathrm{j}} \geq S_{\mathrm{min}} \wedge$$

$$dt(A, t_{\mathrm{j}}, t_{\mathrm{k}}) \leq nParam \wedge dt(C, t_{\mathrm{i}}, t_{\mathrm{j}}) > nParam$$

**Definition 8 (encounter)** . *A moving object encounters a moving context element when it enters its neighbourhood.* This interaction is a step towards the "meeting" interaction. If after the encounter the two objects stay together for some time then the interaction is transformed into a "meeting" one. *For the "encounter" interaction the only condition is that the distance between the target moving object and the moving context element deceases below a predefined nearness parameter.* This is formalised as follows (Note that in the formalisation the last condition ensures that also the context object is moving):

$$\exists t_{\mathrm{i}}, t_{\mathrm{j}} | t_{\mathrm{i}} < t_{\mathrm{j}} \wedge j = i + 1$$

$$d(A, C, t_{\mathrm{i}}) > nParam \wedge d(A, C, t_{\mathrm{j}}) \leq nParam \wedge$$

$$dt(C, t_{\mathrm{i}}, t_{\mathrm{j}}) > nParam$$

**Definition 9 (separating)** . *The moving object separates from the moving context element when it goes outside its neighbourhood.* In other words, the distance between the two objects increases above a specified value of the nearness parameter. This is formalised as follows:

$$\exists t_{\mathrm{i}}, t_{\mathrm{j}} | t_{\mathrm{i}} < t_{\mathrm{j}} \wedge j = i + 1$$
$$d(A, C, t_{\mathrm{i}}) \leq nParam \wedge d(A, C, t_{\mathrm{j}}) > nParam \wedge$$
$$dt(C, t_{\mathrm{i}}, t_{\mathrm{j}}) > nParam$$

**Definition 10 (jointly-moving)** . *A moving object moves jointly with a moving context element when the distance between them does not exceed the value of a predefined nearness parameter throughout their movement.* The *jointly-moving* interaction is formalised as follows:

$$\exists t_{\mathrm{start}}, t_{\mathrm{end}} | t_{\mathrm{start}} < t_{\mathrm{end}}$$
$$\forall t, t_{\mathrm{start}} \leq t \leq t_{\mathrm{end}} : d(A, C, t) \leq nParam \wedge dt(A, t_{\mathrm{start}}, t_{\mathrm{end}}) > nParam$$

**Definition 11 (separately-moving)** . *A moving object moves separately from a moving context element when the distance between them exceeds the value of a predefined neighbourhood parameter throughout their movement.* The *separately-moving* interaction is formalised as follows:

$$\exists t_{\mathrm{start}}, t_{\mathrm{end}} | t_{\mathrm{start}} < t_{\mathrm{end}}$$
$$\forall t, t_{\mathrm{start}} \leq t \leq t_{\mathrm{end}} : d(A, C, t) > nParam \wedge dt(A, t_{\mathrm{start}}, t_{\mathrm{end}}) > nParam \wedge$$
$$dt(C, t_{\mathrm{start}}, t_{\mathrm{end}}) > nParam$$

Case 3 conceptually generalises situations of two or more moving entities. This case covers restricted movement (e.g., vehicle movement following a road network) and less restricted movement (e.g., animal movement in ecology). The implementation of this case may have to take into account situation specific characteristics. For example, while the Euclidean distance can be used to implement this case for animal movement in ecology, the implementation for vehicle movement has to use the distance measured along the road network.

This case (Case 3) of the conceptual model resembles the concept of *dynamic interaction* (Long et al., 2014) in movement ecology and the concept of *relative motion* (Laube, Imfeld and Weibel, 2005). The similarity is that both this case and these concepts indicate how the movement of two or more individuals are related. However, the conceptual model case presented differs from the dynamic interaction methods used in ecology in the way the relation between the movements is computed. In this conceptual model a qualitative approach is adopted while the methods of dynamic

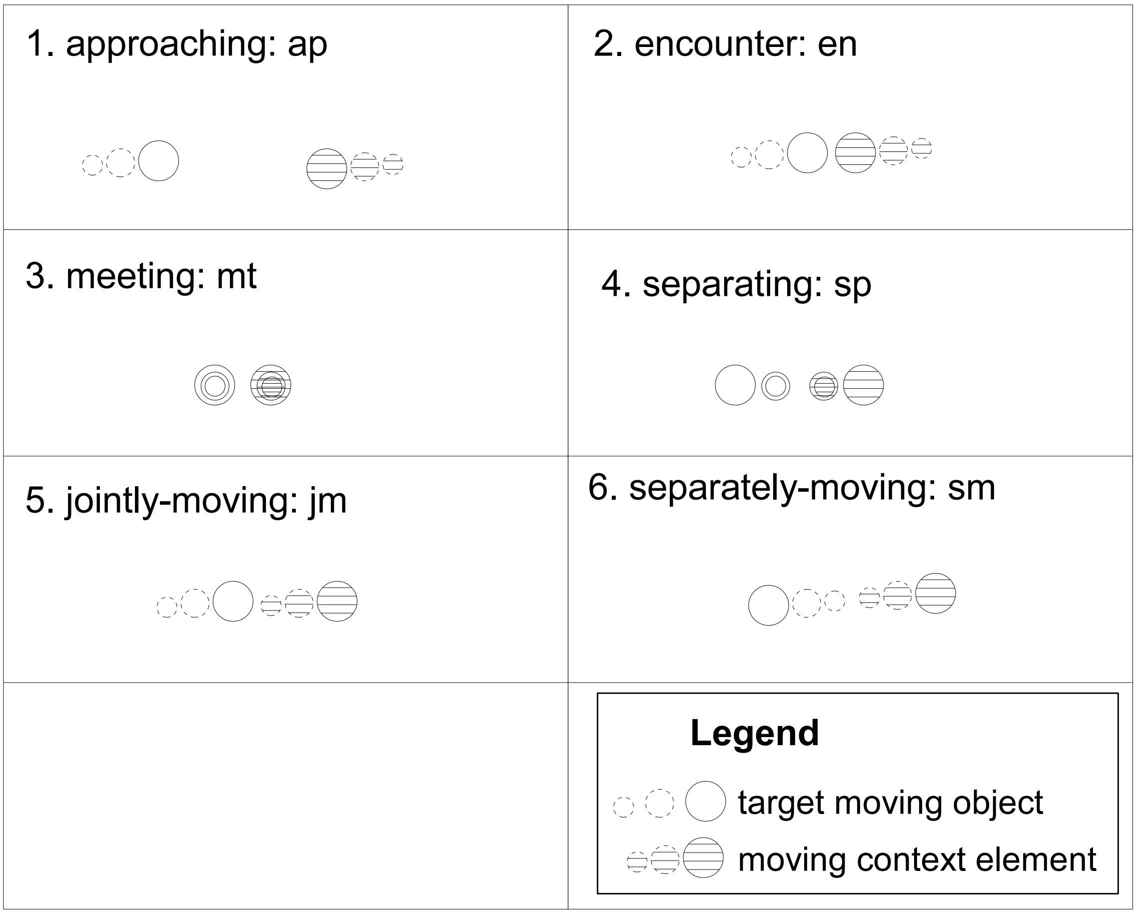



Fig. 3.4 Basic interactions between a moving object and a moving zero-dimensional geographic context

interaction adopt a quantitative approach. For example, the conceptual model proposed in this chapter determines that two moving entities are approaching each other based on the decrease of the distance between them, without caring about how much this decrease is. On the other hand, the dynamic interaction methods in ecology determines that two animals are engaged in attraction based on some quantitative variable (e.g., prox[1], Cr[2] (Long et al., 2014)) which indicates also the degree of attraction.

Although the qualitative approach adopted by the current conceptual model causes a loss of precision, some situations do not need the precision provided by the dynamic interaction methods. Furthermore, the quality of available tracking data may not allow such precision. For example, having tracking data with a lot of missing fixes, it may be enough to determine that two vehicles are approaching each other without trying

[1]Proportion of simultaneous fixes that are spatial proximal
[2]Correlation index

to find their exact locations. In this situation, the locations of the vehicles may not be accurately determined due to the low quality of available tracks. Therefore, case 3 of the proposed conceptual model provides an alternative method to dynamic interaction methods used in ecology. The alternative method is particularly useful in situations of low quality traces (e.g., low sampling rate) and where the requirement of proximity precision can be relaxed.

### 3.2.2 Conceptual neighbourhood graphs of movement interactions

As discussed in section 2.1.3, using conceptual neighbourhood graphs (CNG) to organise qualitative relations between spatial objects helps understanding how they are related and supports reasoning on them. Like other qualitative relations, movement interactions exhibit different degrees of closeness between them, which need to be understood for reasoning on the interactions. Therefore, in order to show how far two interactions are apart from each other, I organise them in conceptual neighbourhood graphs. The conceptual neighbourhood graphs are developed based on the concept of *smooth transformation* (Freksa, 1991). This means that two interactions are conceptual neighbours if a temporal shortening or extension of the first causes a direct transition to the second. The conceptual neighbourhood graphs show between which interactions a gradual change can cause a direct transition, which can in turn support qualitative reasoning on the interactions. Considering the general duration of an interaction, I differentiate *interval interactions* from *instant interactions*. An interval interaction lasts for a time interval while an instant interaction occurs at a certain time moment and does not hold after that time moment.

The conceptual neighbourhood graph of the basic interactions between a moving object and a zero-dimensional context with a fixed location (the first case of interactions presented in section 3.2.1) is depicted in Figure 3.5. A real example of the interactions shown in this CNG is the case of interactions between a car and a traffic light located on the road on which this car is moving. The nodes in the graph are the interactions represented by their abbreviations as introduced in Figure 3.2. An edge between two nodes shows that the interactions represented by the nodes are conceptual neighbours, which means that a direct transition between them is possible. The conceptual neighbourhood Graph shows that *approaching* and *arriving* are conceptual neighbours because if a moving object keeps approaching the context element it will end up by arriving at it and no other interaction will happen in between. The concept of smooth

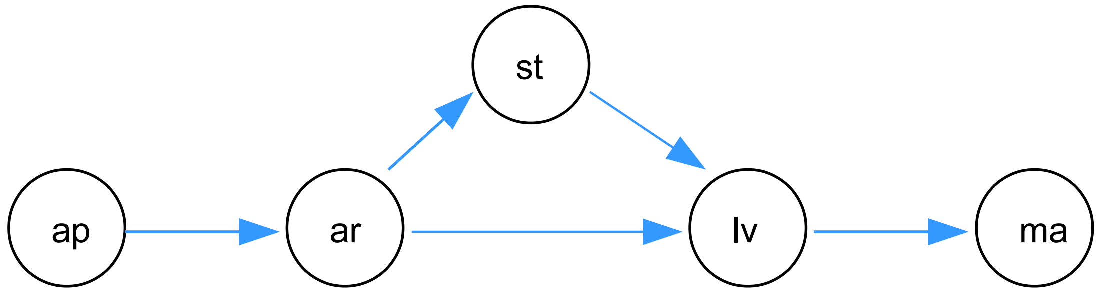



Fig. 3.5 A conceptual neighbourhood graph of interactions with a 0D context having a fixed location

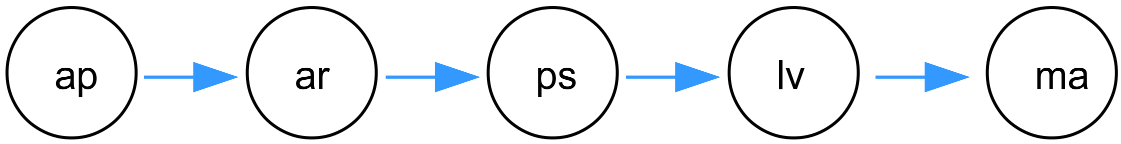



Fig. 3.6 A conceptual neighbourhood graph of interactions with a 1D or 2D context having a fixed location

transformation implies that once the moving object arrives at the context element it can either stop at it or directly leave it. Further, when it has stopped it can cut this stop short to leave the context element. After leaving the context element, the moving object moves away from it as it is shown by the distance between them which keeps increasing. The *arriving* and *leaving* interactions are instant interactions while the others are instant interactions.

In Figure 3.6, the conceptual neighbourhood graph of interactions between a moving object and a 1D or 2D context (e.g., a context element of type “space”) is shown. This CNG organises the movement interactions depicted in Figure 3.3. These interactions can be observed, for example, between a moving car and a road segment. This conceptual neighbourhood graph has a lot of similarities with the CNG shown in Figure 3.5. The similarities are due to the fact that in both cases the context element does not change its location. However, the fact that the context element involved in the case of Figure 3.6 extends over multiple point locations leads to some differences including the new interaction of *passing* the context element.

The interactions between a moving object and a moving zero-dimensional context element depicted in Figure 3.4 are organised in the conceptual neighbourhood graph shown in Figure 3.7. This CNG is very different from the previous two cases. The

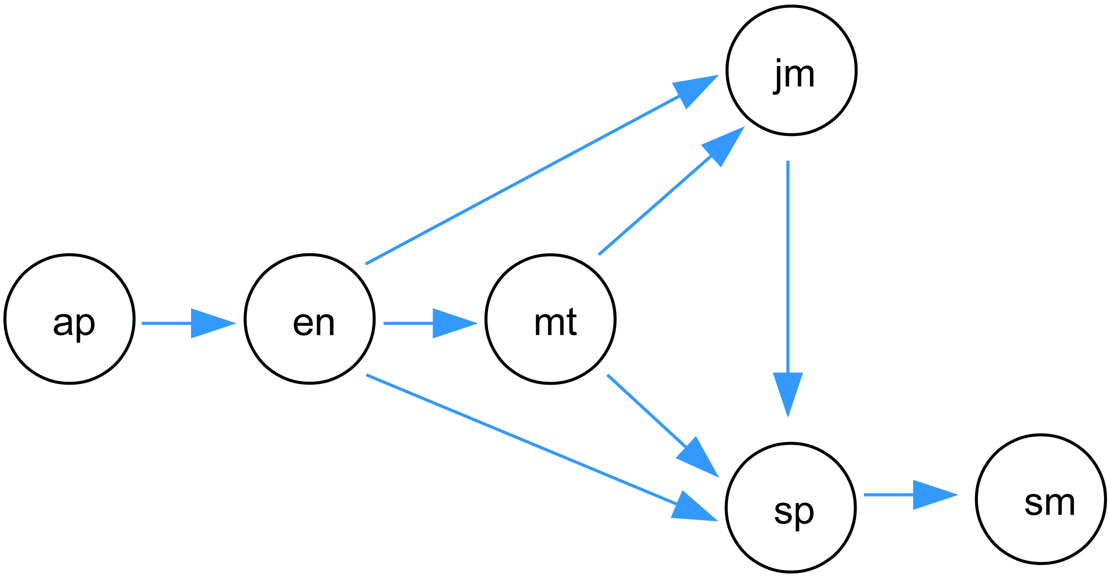



Fig. 3.7 A conceptual neighbourhood graph of interactions with a moving 0D context

differences are mainly due to the fact that the context element also moves. As shown in Figure 3.7, if the target moving object keeps *approaching* the moving context element it ends up by *encountering* it. If the encounter is extended, it leads to the *meeting* interaction. Otherwise, the objects separate or jointly move. The *meeting* interaction is characterised by staying together in the same location for some time. The *meeting* interaction can be cut short leading to *separating* or *jointly moving*. The *separating* interaction occurs at the moment the target moving object leaves the neighbourhood of the context element. This interaction is an instant interaction because immediately after separating, the objects *move separately*. At any time while jointly moving, the objects may separate and start moving separately. The *encounter* interaction is also an instant interaction because immediately after encountering, the objects either stay together in the same location (*meeting*), or move jointly or separates.

## 3.3 Summary

This chapter presented a conceptual model relating the movement of a moving object to its embedding dynamic geographic context. Firstly, the chapter explained the dynamic geographic context to delineate it from the broad concept of context. Then a classification of the geographic context was established and approaches for modelling the different geographic context types proposed. After the clarification of the movement context as considered in this thesis, the chapter introduced the concept of movement interactions. I defined and formalised a small set of movement interactions that abstract the change of spatial relations between the moving object and a context element as the object moves. Finally, I organised the movement interactions into conceptual neighbourhood graphs which can support qualitative reasoning on the interactions. The next chapter develop an analysis framework that exploits the link between the movement and its embedding dynamic geographic context as expressed by the conceptual model presented in this chapter.

# Chapter 4

# A methodological framework for contextualised pattern discovery and analysis

This chapter provides a discussion of the framework proposed for discovering and analysing contextualised movement patterns. The proposed framework exploits the conceptual model of movement context and interactions presented in chapter 3. The chapter starts with a presentation of the general overview of the framework, and then discuss its phases in detail.

## 4.1 Introduction

The proposed framework is based on the KDD (Knowledge Discovery in Database) process (Fayyad et al., 1996) reviewed in section 2.3.1. The steps of the KDD process previously shown in Figure 2.6 are shown again in Figure 4.1 with added labels to show the part corresponding to the proposed framework. In Figure 4.1, the phases before data mining are considered together as the data preparation phase. While the analysis goes through the whole KDD process, the proposed framework corresponds to the steps after the data preparation phase.

The framework (see Figure 4.2) comprises two phases, namely *extraction of movement interactions* and *analysis of movement interactions*, which together correspond to the data mining and pattern interpretation/evaluation phases of the KDD process. The input to the analysis framework is a set of trajectories and corresponding geographic context data that have undergone necessary preparation operations. Both types of data

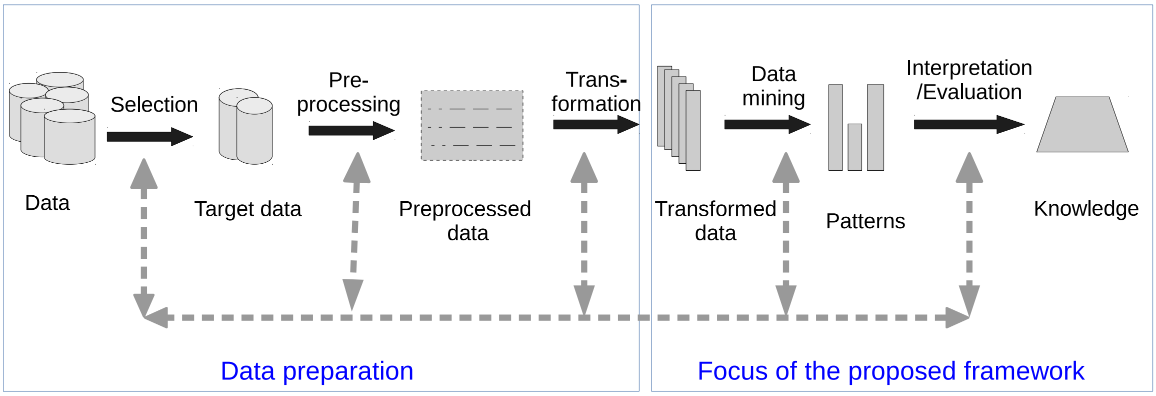



Fig. 4.1 Steps of the KDD process and the focus part of the proposed framework (based on Fayyad et al. (1996))

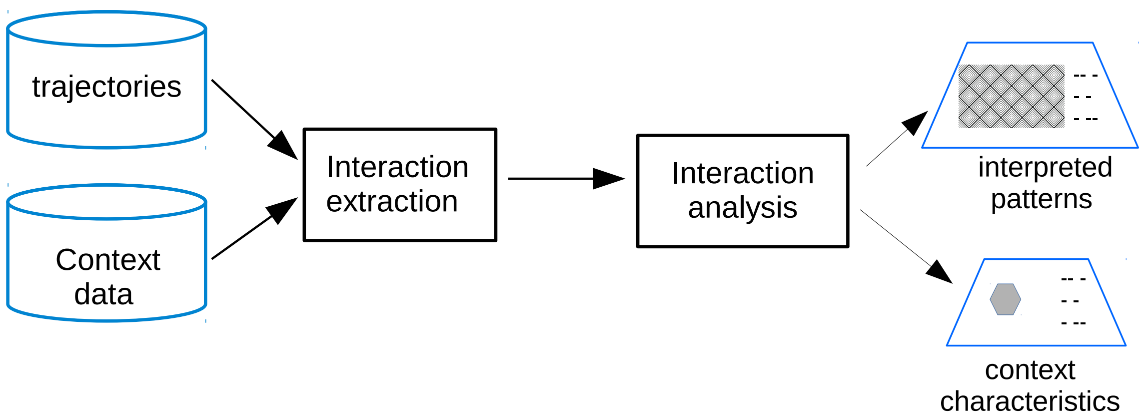



Fig. 4.2 Proposed framework for contextualised pattern discovery and analysis

have been discussed in chapter 2 (trajectories in section 2.2 and geographic context data in section 2.4.4). The output from the proposed analysis framework is either a pattern interpretation or new knowledge about the geographic context or both. The pattern interpretation is a high level description of the movement pattern observed, which can explain why or how the pattern appeared. The new knowledge about the context is a description of some characteristics of the context that were not known at the beginning of analysis.

## 4.2 Extraction of movement interactions

The first step of the proposed analysis framework aims at extracting movement interactions from movement data and associated context data. These interactions represent movement patterns embedding context information. For example, instead of discovering simply a *stop* pattern a *stopping* interaction is extracted (e.g., instead of discovering that a taxi made a stop, it is discovered that it stopped at a traffic light, which can further help to understand why it stopped). In addition to the *stop* pattern, which is purely geometric and represents a part of a trajectory, the *stopping* interaction includes information about the spatial object at which the pattern is found. The extraction of movement interactions is performed using spatial analysis and data mining methods. For example, clustering methods (e.g., density-based clustering) can be applied to identify areas of high concentration of recorded positions of the moving objects which can hint to stops. The movement data or clusters of positions of the moving object can be overlaid with context data for computing distance relations based on which interactions are defined. Other spatial analysis operations (e.g., buffer computation) and spatial relation predicates (e.g., intersection) can be computed to filter out invalid interactions.

There are different data mining methods that can be used in this step. As part of this thesis work, we carried out a review of the literature on generic data mining methods for movement data and their applications. This review, published in (Mazimpaka and Timpf, 2016*a*), has shown that the choice of a method to use depends on the task to be performed, which in turn depends on the application case. Some tasks can be accomplished using one method while others may require a sequence of methods. Furthermore, some methods are related such that a method can be used as standalone to accomplish a task, but also as part of another more complex method. For example, a clustering method can be enough to extract stop patterns (and hence *stopping* interactions) while it can be used as part of a group pattern mining for extracting *jointly-moving* interactions. To concretise this example, we can consider the following analysis cases. In (Jahnke et al., 2017), density-based clustering is used to extract taxi stops as origin and destination hotspots. The stops are then semantically enriched using POIs. With this semantic enrichment we can consider that there are *stopping* interactions between taxis and the POIs. In a second case, Kalnis et al. (2005) extract moving clusters, which can be considered as *jointly-moving* interactions. To this end they use the same clustering as in the first case but they extend it with a method that checks the connectivity of extracted clusters. Therefore, the choice of the data mining method to use depends on the task or the sequence of tasks to be performed for extracting the interactions that occurred.

At the end of interaction extraction, each extracted interaction instance is uniquely identified and has associated attributes such as its start and end times and the identifier of the context element involved in it. These attributes are used for the analysis of interactions in the next step. In addition to static attributes, movement data may also have time-varying (or dynamic) attributes. Some of these dynamic attributes reflect the dynamics of the context element involved. For example, usually a bus has a schedule indicating the time at which it should reach specific locations along a road segment. The movement data of the bus may contain a dynamic attribute indicating the amount of time by which the bus is delayed at any recorded position. In this case, the values of the attribute while traversing a road segment (an instance of the *passing* interaction involving a specific road segment) reflects the "congestion level" attribute of the road segment.

## 4.3 Analysis of interactions

The movement interactions extracted in the preceding phase represent movement patterns embedding context information. In this phase, these interactions are analysed to help understanding the movement patterns that they represent. To this end, statistical methods and further data mining methods are applied. The aim is to quantify the interactions and explore the variation of movement attributes during the interactions to detect and describe any link with the variation of the attributes of the context. The link may be in the form of a correlation, a dependency, or a frequent correspondence between a movement attribute and a context attribute.

In order to take into account the dynamic nature of the context element a temporal window is selected and the analysis of interactions is carried out on time intervals defined based on this temporal window. In the following, some analytical operations are proposed. The proposed analytical operations are based on static and dynamic attributes of the movement and context elements.

1. Extracting summary statistics of attributes in different time intervals and using the statistics to detect correlations, and dependencies.

   For each time interval of the study period, the following summary statistics can be computed for example:

   - the number of interactions involving a specific context element
   - the minimum and maximum values of dynamic attributes during interaction (e.g., maximum speed while approaching the context element)
   - the mean values of dynamic attributes during interaction

- the standard deviations of dynamic attributes during interactions

Based on the summary statistics, the extrema of dynamic attributes can be identified. These are the time intervals with highest or lowest values of the considered attributes. Within the time period covered by the analysis, summary statistics of interactions, dynamic attributes of the movement, and dynamic attributes of the context vary. One way to use the geographic context to support understanding of movement patterns can be to explore these variations. The exploration checks the relation between the variation of a dynamic attribute of the context and the variation of summary statistics of interactions or dynamic attributes of movement. The identification of this relation leads to interpretation of the movement pattern and can also reveal previously unknown characteristics of the context element.

The relation between the variations of two variables can be, for example, that an increase of the first always corresponds to an increase of the second. If both variables are numeric, such dependence or correlation can be checked by computing the *Pearson correlation coefficient* between them. If one variable is nominal, the values of the numeric variable can be classified to check the correspondence between different classes of values and the different values of the nominal variable. In some cases, the analysis may focus on abnormal values (exceptional high or low) of one of the variables; which becomes an *outlier detection* approach. Let `X` and `Y` be two variables, $\bar{x}$ and $\bar{y}$ the mean of `X` and `Y` respectively, $\mathtt{s_x}$ and $\mathtt{s_y}$ the standard deviation of `X` and `Y` respectively.

**The outlier detection approach** involves finding values of the second variable that go out of the bounds established around the normal trend. The relation between the variables is found if there is a regular correspondence between the outliers found and specific values of the other variable. The normal trend can be computed as the mean value. The upper and lower bounds of `Y` are given by:

$Y_{\text{upper}} = \bar{y} + \alpha s_{\text{y}}$ and

$Y_{\text{lower}} = \bar{y} - \alpha s_{\text{y}}$ respectively

The outlier of `Y` is: $\mathrm{y_i} > \mathrm{Y_{upper}}$ or $\mathrm{y_i} < \mathrm{Y_{lower}}$

In the computation of bounds, $\alpha$ is a scaling factor, which controls the uncertainty associated with the length of the time period analysed. The value of the scaling factor defaults to 1 but it should be increased for a long term analysis (say weeks or longer). In other words, a longer period covered by analysis increases the uncertainty about the range of deviation from the normal values. In order to take into account the increased uncertainty, the scaling factor pushes bounds away from the normal trend hence considering a wider range for deviation. The idea of using bounds and their computation come from Bollinger bands (Dobson and Bollinger, 1994) often used in finance. Bollinger bands are bands of standard deviation plotted above and below the moving average of market prices. They help in understanding the market volatility by providing a relative definition of "high" and "low". This allows traders to take advantage of high and low prices. The consideration is that a standard deviation is a good indicator of how the values in a sample fluctuate around the average value. Like in Bollinger bands, the values of a given dynamic attribute of the movement is considered to be high or low (and hence outlier) only with reference to the average value and to how values usually change around this average.

**The Pearson correlation coefficient approach** detects dependence between two variables if the computed correlation coefficient shows that the increase of the value of one tends to correspond to an increase of the value of the other (positive correlation) or the increase of the value of one tends to corresponds to a decrease of the value of the other (negative correlation) (Ross, 2014). The correlation coefficient between `X` and `Y` is given by:

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

The Pearson correlation coefficient is appropriate when the two variables are numeric. In case any of the variables is nominal the outlier detection approach can be used instead.

2. Sequence analysis

   Another analysis operation is the analysis of sequences of interactions. As shown by the conceptual neighbourhood graphs presented in section 3.2.2, movement interactions are related through possible transitions between them. This implies that, given an interaction in which a moving object is participating, it is possible to predict the next interaction in which the object is likely to participate. Therefore, the comparison of the sequence of interactions in which a trajectory is involved to expected sequences can reveal abnormalities. Sequence analysis aims at detecting these anomalies and possibly tracing them back from the values of some attributes of the context element through a qualitative reasoning on the interactions.

## 4.4 Summary

In this chapter an analysis framework for trajectories and associated dynamic geographic context has been discussed. The framework exploits the conceptual framework presented in chapter 3, which relates the movement to the context. The framework comprises two steps, namely interaction extraction and interaction analysis, which are together aimed at discovering and interpreting movement patterns. Some methods that can be used under each step have been presented. In particular, under interaction analysis different analysis operations have been proposed. They include the detection of correlation and dependence between dynamic attributes of movement, number of interactions, and the dynamic attributes of the context. Also sequence analysis of interactions has been proposed for movement prediction. The next chapter evaluates the applicability of the proposed analysis framework using real data.

# Chapter 5

# Experiments

This chapter provides a summary of experiments executed to evaluate the applicability of the proposed framework. As discussed on modelling the geographic context (section 3.1), the geographic context can be modelled in many different ways. Moreover, its dynamics are based on the change of its location, its lifespan, time-varying attributes, or change of extent. The experiments presented in this chapter address different aspects regarding the dynamics of the geographic context and the analysis framework proposed. For each experiment, the data used are briefly described and then the analysis framework applied.

## 5.1 Experiment 1 – temporal dynamics of a dynamic geographic context

The main objective of this experiment is to demonstrate the applicability of the proposed analysis framework on a geographic context which has a lifespan shorter than the time period covered by the analysis. Within the period covered by the analysis the context element comes into existence and ceases to exist. This defines three phases "before", "during", and "after" such that the beginning and the end of the lifespan may need to be discovered. Once the beginning and end of the lifespan are known, their characteristics as well as effect on the movement may need to be determined. Therefore, the main focus is on the time dimension related issue: whether the times of appearance and disappearance of the context element can be discovered from movement data and whether different phases of the lifespan of the context element can be characterised in terms of movement patterns. This section (5.1) is based on (Mazimpaka and Timpf, 2017).

Social events have been selected as context elements because they are a good example of the type of context dealt with in this experiment. The lifespan of a social event is shorter than the time period covered by the analysis. During its time of existence, a social event goes through different phases which can be reflected differently on human mobility pattern. Alternatively, other types of spatial events that occur in the geographic space of movement such as a roadwork and road accident could be considered. The aim of the case studies in this experiment is to support understanding of the dynamics associated with big events and human mobility. That is, given mobility data in the neighbourhood of a venue known to host big social events, can we detect the occurrence of events at the venue? If an event is detected can we learn from the movement data its characteristics which explain in turn the movement patterns observed?

### 5.1.1 Datasets

The movement data analysed in this experiment are trajectories of buses in Dublin. The dataset is a subset of the Dublin bus GPS dataset[1] from the Dublin City Council's traffic control. Each bus produces a record about its location and status every 20 seconds on average. The record includes a timestamp, latitude and longitude of the location, the bus line ID, the vehicle ID, the journey pattern (an indication of the direction), an identifier of the closest bus stop, the delay (number of seconds for which the bus is behind the schedule, which is negative if the bus is ahead of schedule), whether the bus is at a bus stop, and whether the bus is in a congestion. Specifically, this experiment analyses trajectories of buses following bus lines 4 and 44.

The data about context elements include the locations of bus stops along the routes used by the two bus lines, and the locations of the Aviva stadium and the National Concert Hall where events take place. The study area is shown in Figure 5.1. The left part of the figure (a) shows the route used by bus line 4. This route has a length of approximately 20 km and includes 65 bus stops in one and 61 in the other direction. The right part of the figure (b) shows the route used by bus line 44, which includes 80 bus stops in one and 76 in the other direction. Figure 5.1 also shows the location of the Aviva stadium and a neighbourhood of 650 meters around it (left part), and the location of the National concert hall and a neighbourhood of 350 meters around it (right part). The sizes of these neighbourhoods are selected to enclose all bus stops that serve directly the venues such that a passenger dropped there may not take any

[1]https://data.dublinked.ie/dataset/dublin-bus-gps-sample-data-from-dublin-city-council-insight-project

other bus to reach the venue. For each bus stop the data include a unique identifier, a name, GPS coordinates and the distance at which it is located from the beginning of the route.


Fig. 5.1 Location of context elements: a) line 4 bus stops and Aviva stadium, and b) line 44 bus stops and the National Concert hall

As ground truth, we retrieved from the Website of the National Police Service of Ireland[2] and Wikipedia information about the occurrence of big events in the Aviva stadium in the time period covered by the movement data. This event occurrence information is shown in Table 5.1. Likewise, we retrieved from an event archive Website[3]

[2]http://www.garda.ie/News/default.aspx
[3]https://issuu.com/nationalconcerthall/docs/sept-nov2012calendar

information about the concerts that took place in the National Concert hall in the period covered by the movement data.

| Date | Event | Planned stiles opening time | Planned start time | Planned end time | Actual end time | Number of attendees |
|---|---|---|---|---|---|---|
| 10/11/2012 | Rugby match (Ireland V South Africa) | 16:00 | 17:30 | 19:00 | 19:20 | 49,781 |
| 14/11/2012 | Football match (Ireland V Greece) | 18:15 | 19.45 | 21:45 | 21:42 | 16,256 |
| 24/11/2012 | Rugby match (Ireland V Argentina) | 12:30 | 14:00 | 15:40 | 15:49 | 43,406 |

Table 5.1 Occurrences of big events in the stadium during the study period

We prepared the bus movement data through the following pre-processing operations. We performed time format and coordinate system transformations. Next, we discarded unrealistic GPS points. These were identified as points to which the travel from preceding point was found to have been done with a speed higher than 50 km/h. We recomputed, for each GPS point, the nearest bus stop because we found wrongly assigned nearest bus stops. We discarded GPS points that were causing oscillation back and forth along the movement of the bus. These were identified by comparing the distances between three consecutive points. In the following step individual bus journeys were identified and assigned unique identifiers. We then discarded journeys with a large gap in space or in time between any two consecutive recorded positions. Likewise, incomplete journeys on their start or end were identified by checking the closest bus stop associated with their first and last recorded positions respectively and discarded. The final clean bus data for the bus line 4 contained 2,249 journeys made on 28 days between November 2012 and January 2013. This period includes 3 days (10th, 14th, and 24th November 2012) with big events in the Aviva stadium. The same cleaning operations were carried out on the data of bus line 44.

### 5.1.2 Extraction of movement interactions

After the data preparation we extracted interactions between the bus and bus stops, and between the bus and potential events at the selected venues. Bus stops constitute

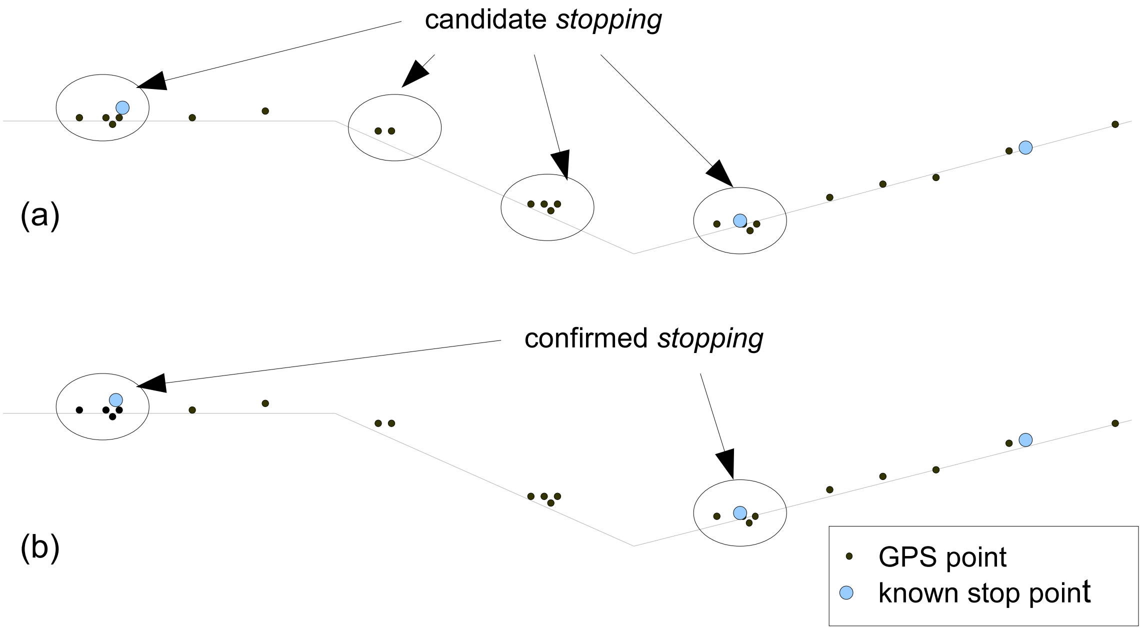



Fig. 5.2 Extraction of stopping interactions

a static geographic context while events constitute a dynamic context. In this case study, the intended analysis needs only three types of interactions: *stopping* (at a bus stop or at an event), *approaching* an event, and *moving-away* from an event.

Although a bus stops at pre-defined locations (called bus stops) along its route, in most cases the bus stops only when there are passengers who want to enter or exit the bus. Therefore, it cannot be assumed that each time a bus reaches a bus stop there is a *stopping* interaction. We computed the interactions as actual stop at the pre-defined stop point. From the definition of a *stopping* interaction given in section 3.2.1, the task is to check if the initially moving bus stays in the neighbourhood of a bus stop for at least a specified amount of time. This can be implemented in different ways. Our implementation works in two steps like the approach used in (Palma et al., 2008): detecting a stop and checking that the detected stop is in the neighbourhood of a bus stop. Considering that staying in a location for some time leads to densely located GPS points, we detected a stop using a density-based clustering. The second step checked each candidate *stopping* detected in the first step to see if it is located within a pre-defined buffer around a known bus top. Candidate *stoppings* found outside the buffer were discarded. This process is illustrated in Figure 5.2.

Since the detection of event occurrence is one of the objectives of the analysis, the *approaching* and *moving-away* interactions were extracted with reference to a potential event represented by the event venue. It means that the event occurrence was to be detected by identifying the interactions (*approaching* and *moving-away*) that relate to actual event. For each journey, the extraction of *approaching* and *moving-away* interactions aims at identifying the section where the moving object is approaching the event and the section where it is moving away from the event. From each route end we identified the known stop point closest to the event location taking into account the road segment providing access to the event location from the route. These two stop points form reference points in the two directions of the route for identifying journey sections. Based on the journey direction and the variation of the distance along the route to the respective reference point we marked the journey points. When the distance decreased, the point was marked to be on an *approaching* section and when it increased, it was marked to be on a *moving-away* section. When the distance did not change marking was deferred until following change. This process followed the definitions of *approaching* and *moving-away* given in section 3.2.1.

After the extraction of interactions each journey position is marked as whether it is part of a *stopping* interaction or not. In case it is part of a *stopping* interaction, the position is also marked as whether it is on the *approaching* or *moving-away* section of the journey. Figure 5.3 shows an example of interactions extracted from a journey segment. The neighbourhood of a bus stop (`nParam`) was defined by a buffer of 20 meters. The value was selected based on the fact that a GPS point could be recorded up to 20 meters away from the actual position due to the GPS accuracy of 20 meters. The neighbourhood distance and the minimum number of points for density-based clustering were set to 20 meters and 2 respectively based on experimentation with different values and the need to detect even short stops. With the same reasoning, the minimum stop duration ($\mathtt{S}_{\mathtt{min}}$) was set to 10 seconds.

### 5.1.3 Analysis of movement interactions

The analysis of extracted interactions was performed using the analytical operations proposed in chapter 4. Specifically, we extracted summary statistics on the interactions and used them to detect outliers. The assumption is that different phases of an event affect the movement patterns of buses differently and this is reflected in the interactions. In other words, the movement pattern observed when there is no event is different from the one observed when an event is about to start and different from the one observed when it has just ended. So we wanted to learn from the interactions about events (occurrence, start time, end time ...). The challenge was to find an appropriate way of quantifying the interactions over the period studied to reveal these dynamics of events.

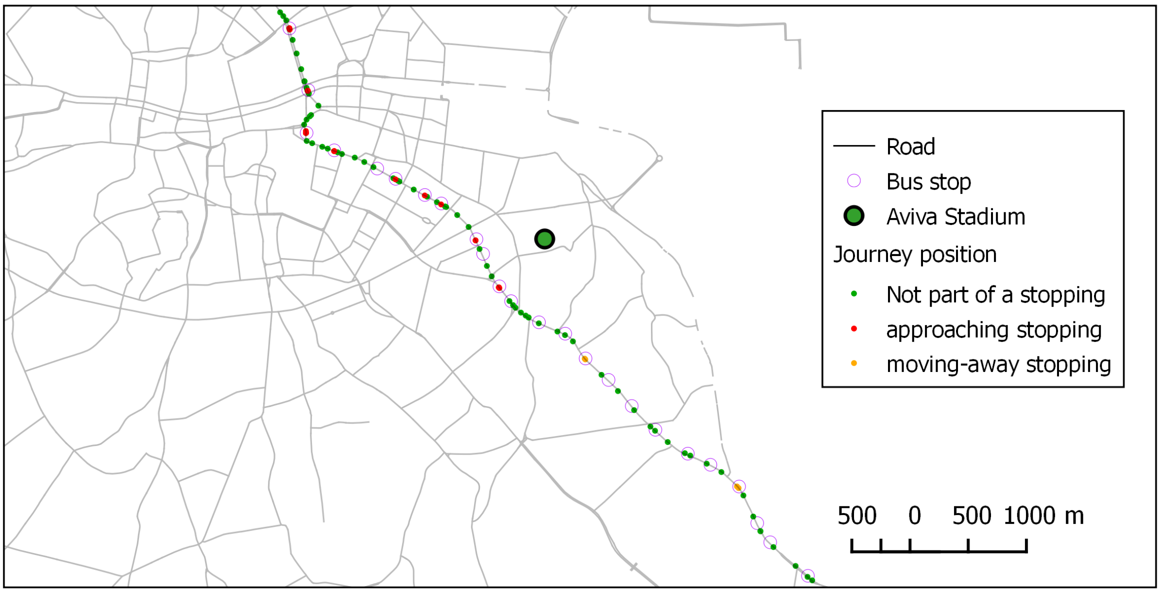



Fig. 5.3 Interactions extracted on a journey segment

We proposed to quantify interactions at multiple temporal granularity levels to detect global and local patterns from the interactions. It means that we had to define multiple time windows based on which we could aggregate the interactions. At the larger time window we detect outliers that correspond to a change in the general mobility pattern, and a possible indication of the occurrence of event. At the smaller time window, we refine the rough estimation from larger time window by focusing on local patterns over a shorter time period. To this end, we considered a 1-hour interval and computed summary statistics on interactions in each 1-hour interval between 6:00 and 23:00 on each day of the study period.

The assumption that a bus stops at the pre-defined stop points generally when there are passengers that want to enter or exit it leads to a positive correlation between the number of stoppings and the situation at the event venue. Likewise, the duration of a *stopping* correlates positively with the number of passengers entering and/or exiting the bus such that in case of an event attracting a lot of people the average *stopping* duration is likely to increase. With these considerations, we explore the variation of the number of stoppings to infer the dynamics of events at the selected venue.

We defined the following variables that were aggregated in the time intervals studied:

- The *number of stoppings near the venue (P)*: to capture all *stoppings* that can be directly attributed to the venue irrespective of the bus line. They are *stoppings* occurring in the venue neighbourhood (see Figure 5.1).

- The *stoppings balance (V)*: The difference of the proportions of *stoppings* while approaching the venue and *stoppings* while moving away.

For each of these two variables, we defined a corresponding normal variable determined in experimental conditions where no event was occurring:

- The *normal number of stoppings near the venue (Q)*
- The *normal stoppings balance (W).*

In the same conditions the normal variables are determined, corresponding standard deviations ($S_Q$, $S_W$) are also determined such that we obtain the bounds of $P$ and $V$ as follows:

- Upper bound of $P$: $UP_i = Q_i + \alpha S_Q$
- Upper bound of $V$: $UV_i = W_i + \alpha S_w$
- Lower bound of $V$: $LV_i = W_i - \alpha S_w$

where $\alpha$ is a scaling factor which defaults to 1. We recall from section 4.3 that this scaling factor controls the uncertainty associated with the length of the time period analysed. That is; the longer the time period, the more uncertain the location of the bounds. To cater for this uncertainty we push the bounds away from the normal to consider a wider range of deviation from normal values. This is done by increasing the scaling factor above 1.

The interpretation of the relations between these variables is as follows. The normal variables indicate the mobility pattern in normal conditions; i.e., without influence of an event which attracts unusual number of passengers. The outliers of `P` and `V`; i.e., values beyond their corresponding bounds (`UP,UV,LV`), indicate an unusual mobility pattern caused by an event at the event venue. The characteristics of the event (e.g., starting phase, ending phase) that cause the unusual mobility are determined from the type of outlier (i.e., below lower bound vs. above upper bound). Specifically, a higher value of `V` above the upper bound `UV` means exceptionally more *approaching stoppings* than *moving-away stoppings*, which corresponds to arrival of event attendees. Conversely, a lower value of `V` below the lower bound `LV` means exceptionally more *moving-away stoppings* than *approaching stoppings*, which corresponds to departure of event attendees. Figure 5.4 shows the variation of `P` and its related variables `Q` and `UP` on a sample day without event while Figure 5.5 shows the variation of `V` and its related variables `W, UV`, and `LV` on the same day.

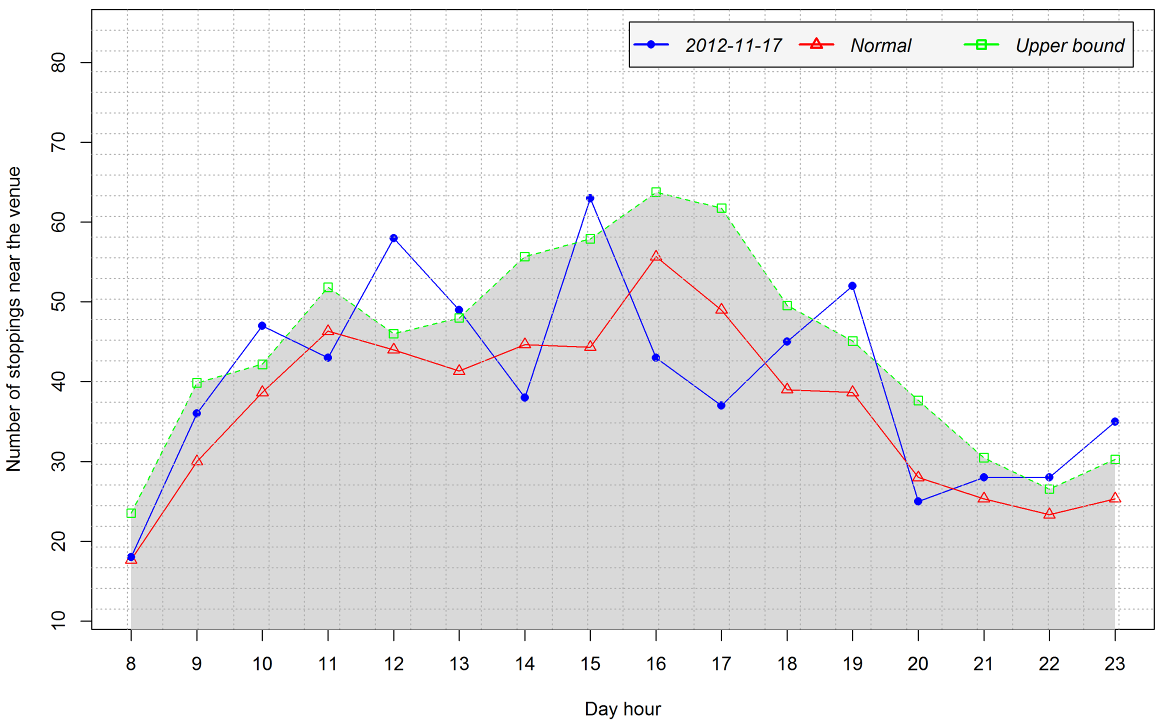



Fig. 5.4 Temporal variation of the *number of stoppings near the venue* (`P`), its normal value (`Q`) and, and its upper bound (`UP`) on a day without event

From the variation of P with respect to its related variables we get candidate event indicators that need to be confirmed using the variation of `V` with respect to its related variables. That is; the peaks of `P` that exceed the corresponding upper bound values are candidate event delimiters in time (arrival and departure) that can be confirmed by checking the variation of *stoppings* balance between *approaching* and *moving-away*. From the example shown in Figure 5.4, we have four candidate event delimiters: 10:00, 12:00, 15:00, and 19:00. For each candidate we take the interval from the last day hour (before it) at which `P` was below the upper bound to the first day hour (after the candidate) at which `P` was below the upper bound. This interval allows us to take into account uncertainties due to aggregating data into 1-hour intervals; hence we call it the 'uncertainty interval'.

We search for a peak in the variation of `V` (see Figure 5.5) within each uncertainty interval. We distinguish two types of peaks. If the value at a certain time in the interval is higher than all preceding and following values in the interval we have a 'positive peak'. If the value at a certain time in the interval is less than all preceding and following values in the interval we have 'a negative peak'. The time corresponding to a positive peak is a candidate arrival time because it corresponds to an exceptionally

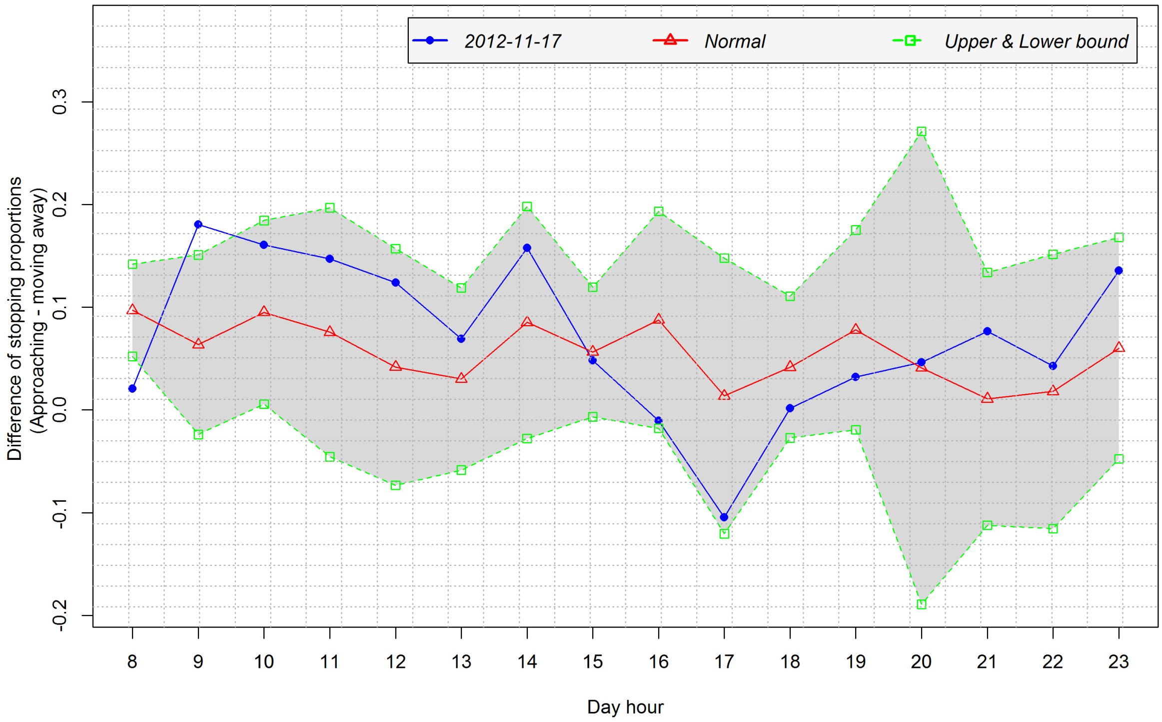



Fig. 5.5 Temporal variation of the difference of *stoppings* proportions between *approaching* and *moving-away* (`V`), its normal value (`W`) and its upper and lower bounds (`UV, LV`) on a day without event

high number of *approaching stoppings*. On the other hand, the time corresponding to a 'negative peak' is a candidate departure time because it corresponds to an exceptionally high number of *moving-away stoppings*. The search for peaks in the variation of `V` (see Figure 5.5) within the four uncertainty intervals obtained from the variation of `P` (see Figure 5.4) results in no peak, which means that the four candidate event delimiters were minor abnormalities in the mobility pattern near the stadium. No event occurred because the global mobility pattern shows no exceptional change (no outliers).

While Figure 5.4 and Figure 5.5 show the variation of the variables on a day without event, Figure 5.6 and Figure 5.7 show the variation of the same variables on a day with event. As seen in Figure 5.6, there are four candidate event delimiters: 9:00, 13:00, 16:00, and 20:00 located in the uncertainty intervals: 8:00 to 11:00, 11:00 to 14:00, 15:00 to 18:00, and 19:00 to 21:00 respectively. The search for peaks in these intervals from the data presented in Figure 5.7 finds a positive peak in the interval containing 13:00 and a negative peak in the interval containing 16:00. Therefore, an event is confirmed to have occurred and the peaks at 13:00 and 14:00 indicating arrival and departure of attendees respectively are taken as the event start and end times respectively.

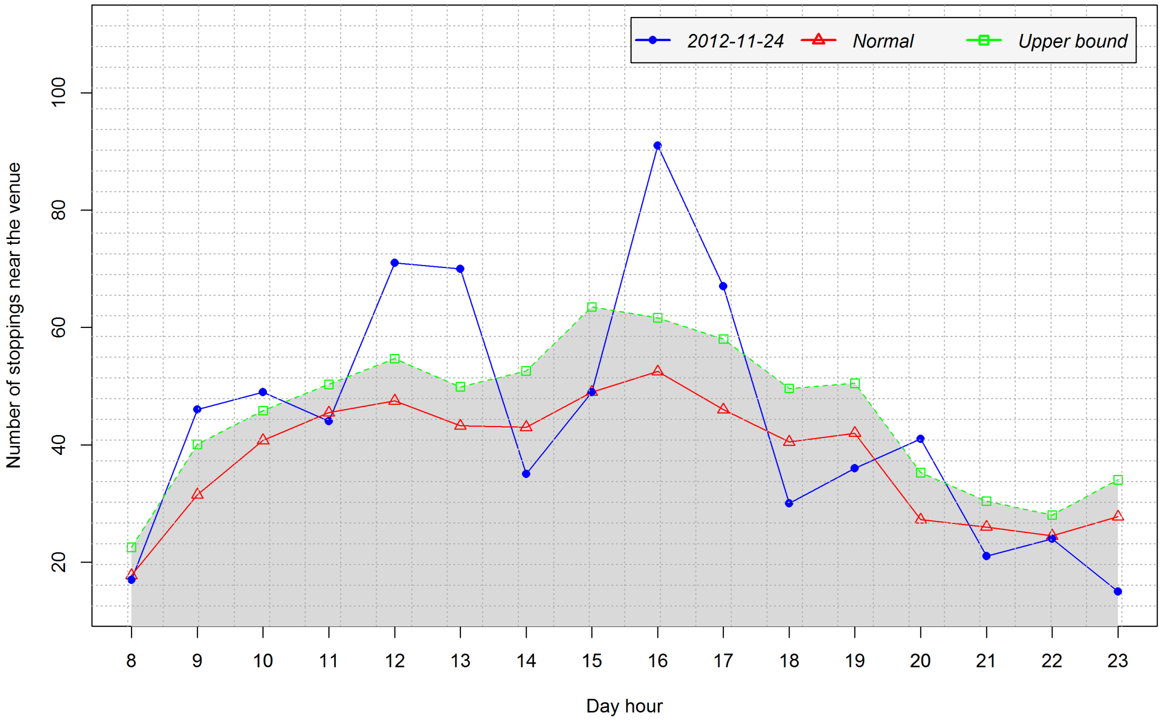



Fig. 5.6 Temporal variation of the *number of stoppings near the venue* (`P`), its normal value (`Q`) and, and its upper bound (`UP`) on a day with event

In order to refine the estimation of event start and end times and to explore the temporal patterns of arrival and departure of event attendees, we perform a local analysis of *stopping* interactions at a finer temporal granularity level. To this end, we explore the variation of the *number of stoppings near the venue* using a temporal window of 15 minutes for aggregating the variables. The focus is on the uncertainty intervals containing the estimated event start and end times. Figure 5.8 shows the temporal variation of the variables within the uncertainty intervals 11:00 to 14:00 and 15:00 to 18:00. This analysis allows us to refine the answer to the question of estimating the start and end times of the event assuming that the highest peak corresponds to the start or end of the event. From Figure 5.8(a) we see that the start time estimated in the previous step to be 13:00 is refined to be around 13:30. Similarly, Figure 5.8(b) shows a refinement of the end time from 16:00 to around 16:15.

The analysis further shows the temporal patterns of arrival and departure of event attendees. The temporal pattern of arrival presented in Figure 5.8(a) shows that some event attendees have been arriving earlier before the event start time as shown by the shorter peaks that exceed the upper bound between 11:00 and 13:00. After the start of the event (approximately after 15 minutes) the number of *stoppings* at the

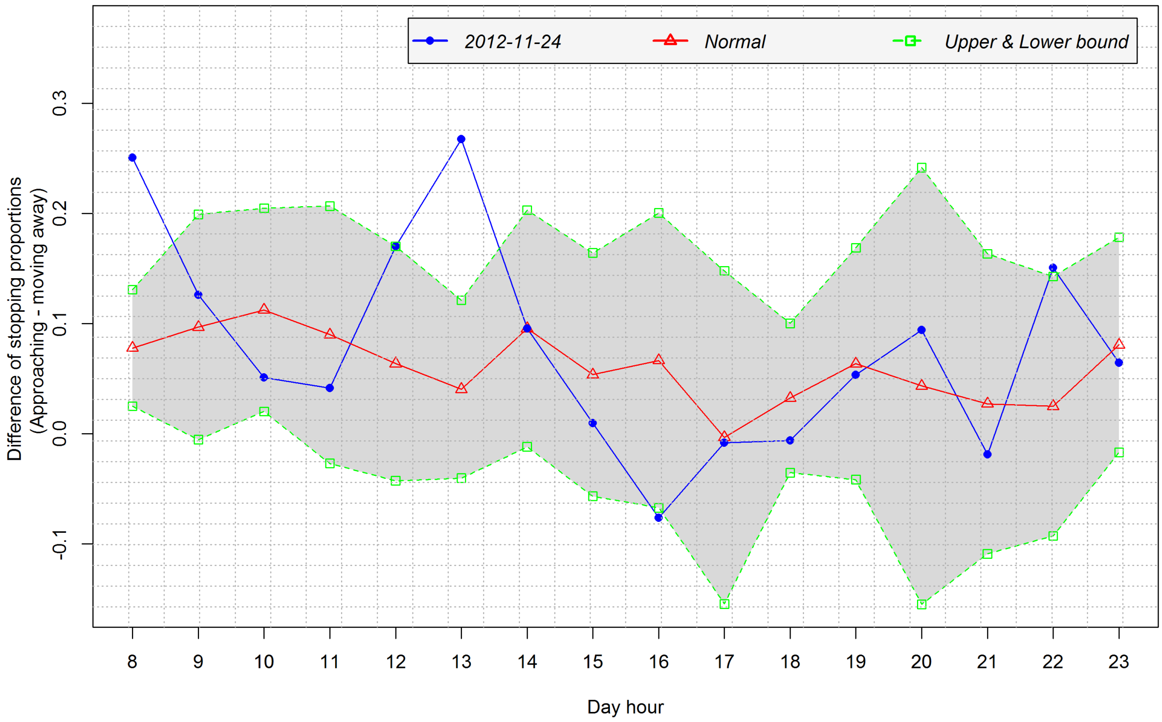



Fig. 5.7 Temporal variation of the difference of *stoppings* proportions between *approaching* and *moving-away* (`V`), its normal value (`W`) and its upper and lower bounds (`UV, LV`) on a day with event

venue sharply dropped below the upper bound becoming almost normal. This suggests that in general event attendees arrived on time. The temporal pattern of departure shown in Figure 5.8(b) suggests that it has taken less than 30 minutes after the end of the event for the *stoppings* near the venue to return to normal; meaning that event attendees did not spend much time at the venue after the end of the event.

Different big events may show different temporal patterns of arrival and departure of attendees. For example, unlike the attendees of the event on 24/11/2012 who arrived on time and departed as soon as the event ended (see Figure 5.8), the attendees of the event on 14/11/2012 kept arriving after the start of the event as shown by the *number of stoppings near the venue* that remained above the upper bound for some time after the peak (see Figure 5.9(a)). Attendees of the latter event also departed progressively as shown by the *number of stoppings near the venue*, which remained above the upper bound during a long time interval after the peak (see Figure 5.9(b)).

### Case of a medium scale event as a dynamic geographic context

In this section we present the results of analysis on a case involving a medium scale event so that we can compare with the case involving a large scale event. Compared

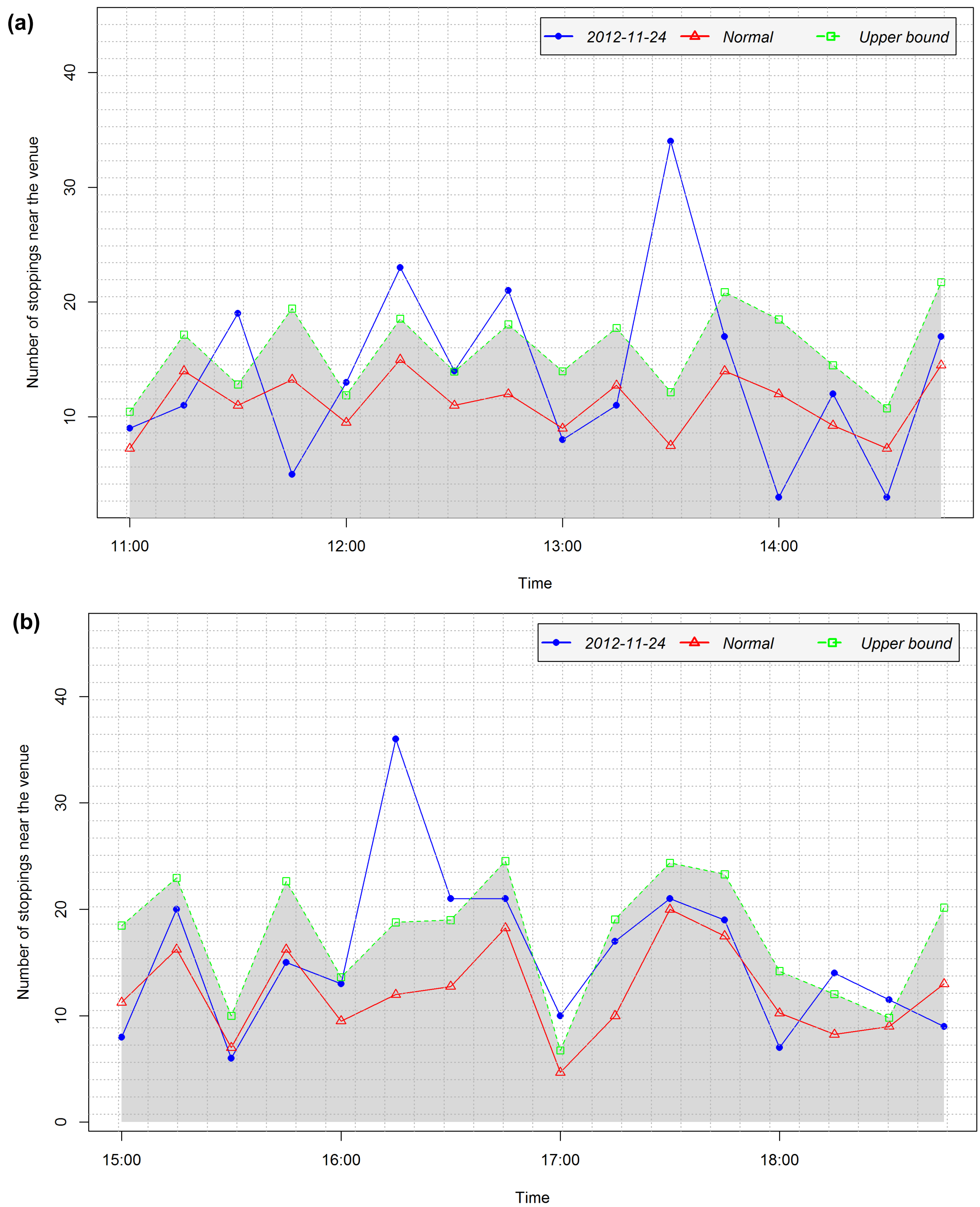



Fig. 5.8 Temporal variation of the *number of stoppings near the venue* (`P`), its normal value (`Q`) and, and its upper bound (`UP`) during the period around arrival (a) and departure (b) times on 24/11/2012

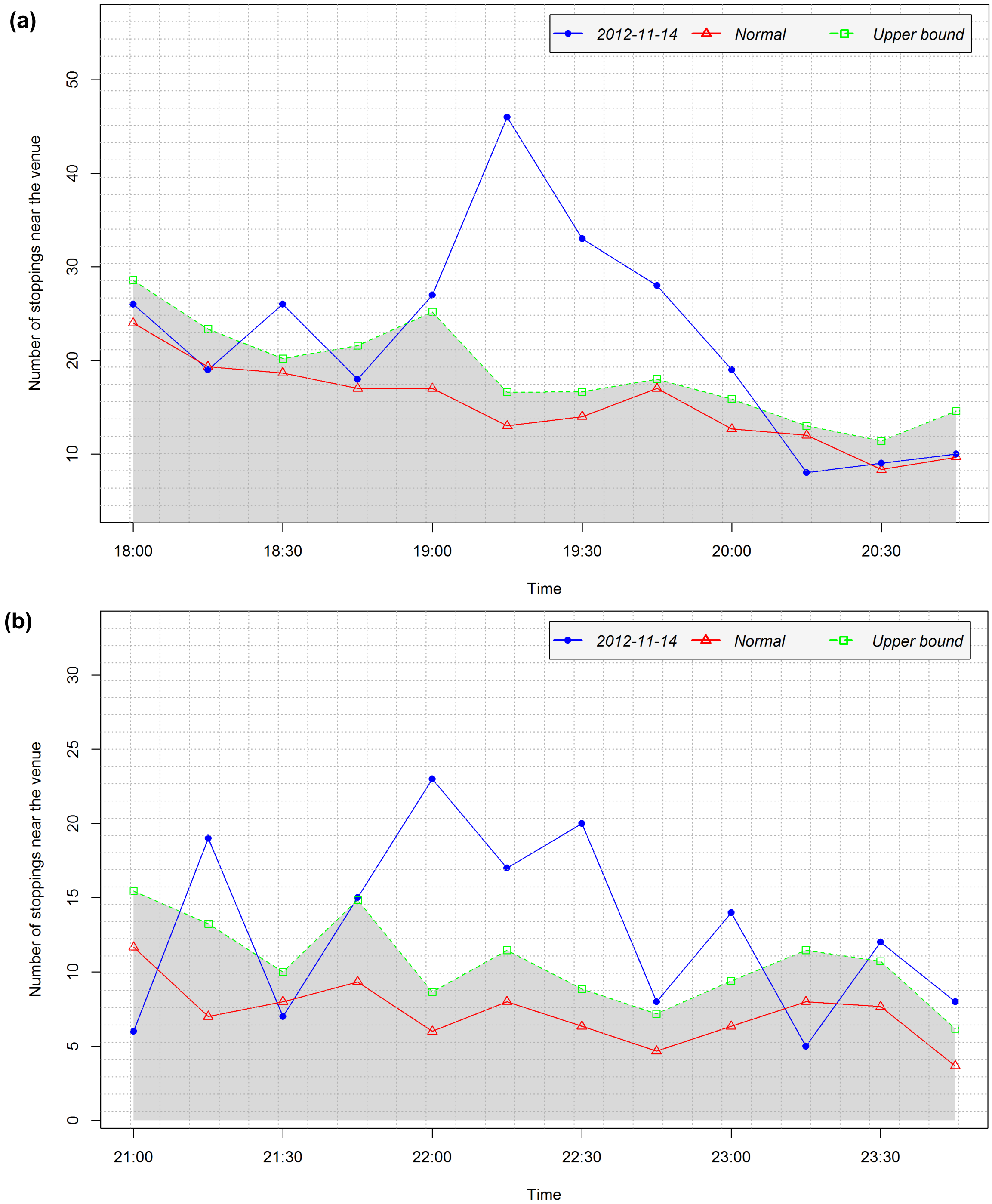



Fig. 5.9 Temporal variation of the *number of stoppings near the venue* (`P`), its normal value (`Q`) and, and its upper bound (`UP`) during the period around arrival (a) and departure (b) times on 14/11/2012

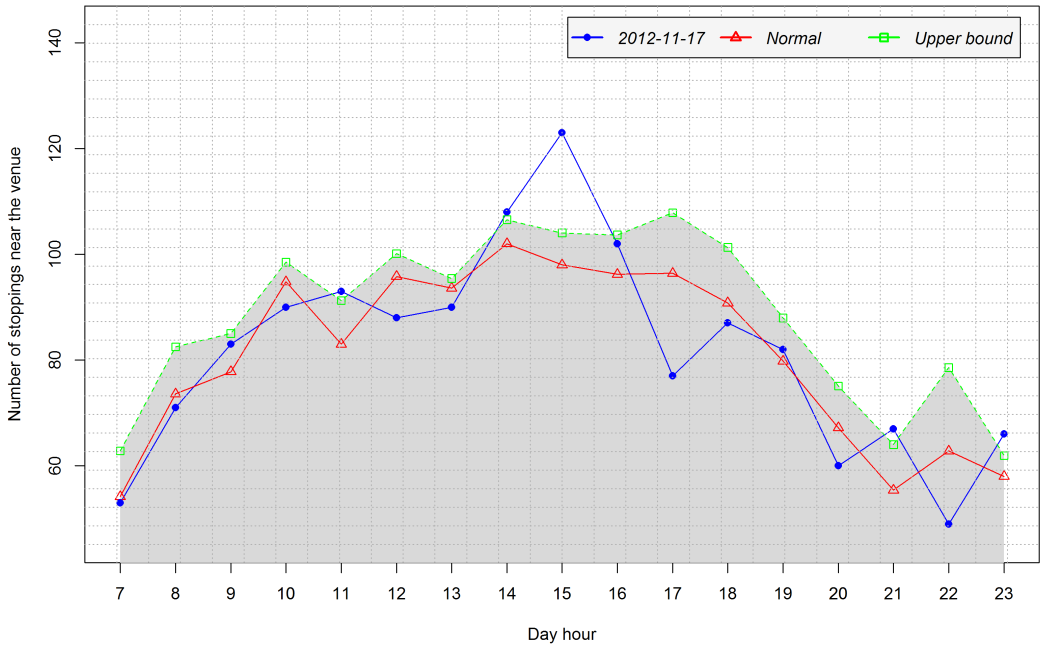



Fig. 5.10 Variation of the number of bus *stoppings* at the National Concert hall on a day with event

to a large scale social event such as the one already presented, a medium scale social event attracts fewer people. We carried out the second case study following the same steps explained in detail on the case of large scale event. For this case of medium scale events (see the study area in Figure 5.1(b)), events were organised at the same hour of the same day every week during the period studied. This led to difficulty in finding a model of the normal condition. Therefore, we analysed only two days (17th and 28th November 2012) for which a model of the normal mobility could be obtained. The analysis results for the 17th November 2012, on which an event was organised at a different hour (14:30) compared to other days (20:30), are presented here as an example. Figure 5.10 shows the variation of the number of bus *stoppings* in the neighbourhood of the concert hall while Figure 30 shows the variation of the balance between the number of *approaching stoppings* and *moving-away stoppings*. A candidate event occurrence was found at 15:00 in the uncertainty interval of 13:00 to 16:00 (see the peak in Figure 5.10). This candidate was confirmed to be the start of an event (see the positive peak in the corresponding uncertainty interval in Figure 5.11). Through the analysis of the uncertainty interval at a finer granularity level, the event start time was refined to be around 15:45 (see Figure 5.12).

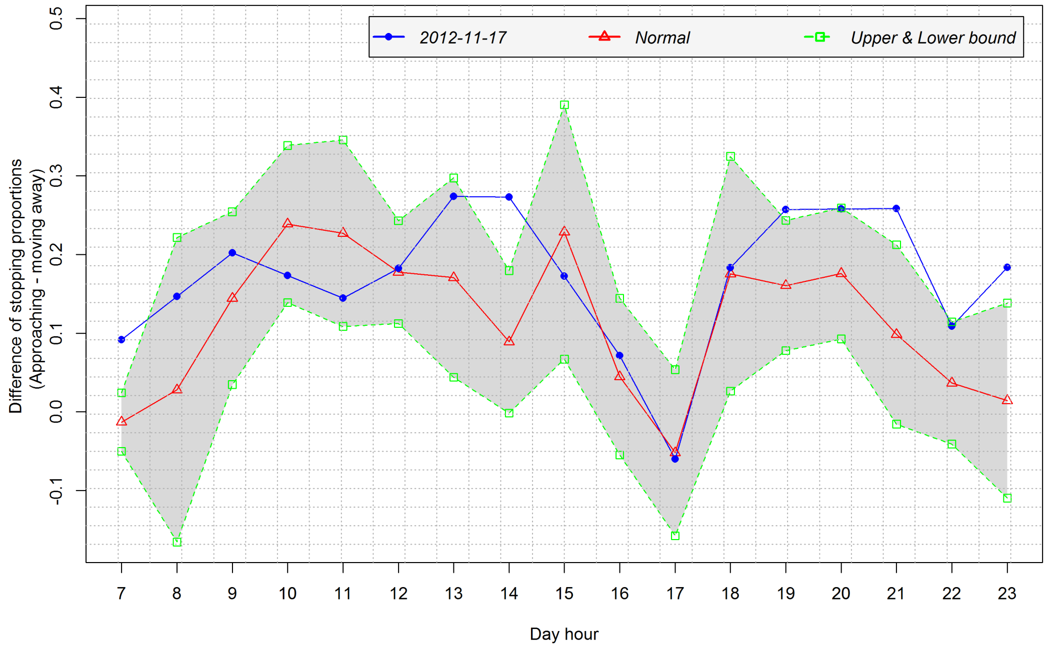



Fig. 5.11 The variation of the balance of *stoppings* while approaching the event venue and *stoppings* while moving away on a day with event

### 5.1.4 Summary and discussion of results of experiment 1

In this experiment, the proposed framework for context-aware movement data analysis has been applied to cases involving a dynamic geographic context having a bounded lifespan. The context element comes into existence and ceases to exist within the period covered by the analysis. Typical such context elements are spatial events. Specifically, social events attracting a considerable number of people have this characteristic and have been selected as case studies. Three types of movement interactions have been studied: *stopping, approaching* and *moving-away*.

The interactions have been extracted from the movement data of buses running on two bus lines (4 and 44) in the city of Dublin. Line 4 bus data were analysed with the context of large scale events that take place in the Aviva stadium while line 44 bus data were analysed with the context of medium scale events that take place in the National Concert hall. The interactions were extracted using data mining and spatial analysis methods. *Stopping* interactions were extracted using clustering, buffer creation and topological relations checking. *Approaching* and *moving-away* interactions were extracted using distance measurement. Some *stopping* interactions relate the bus to pre-defined bus stop locations while others relate the bus to a potential event

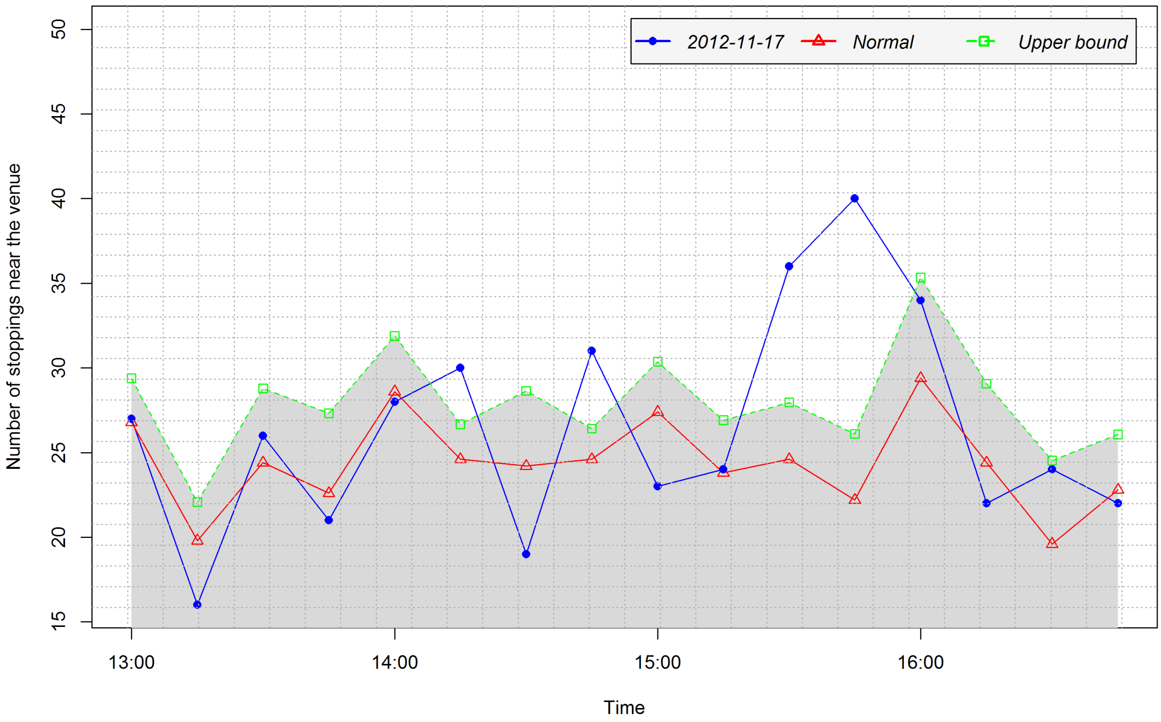



Fig. 5.12 The variation of the number of bus *stoppings* at the National Concert hall during the period around the arrival time of event attendees

(i.e., a location where an event is taking place or where it is expected to take place). *Approaching* and *moving-away* interactions relate the bus to a potential event.

The extracted interactions were analysed to interpret the bus movement patterns that they represent. The objective was to show the relation between the dynamics of events and the movement pattern of the bus. Since the event is not always present, the analysis had to detect it first. To this end, we performed the analysis at multiple granularity levels on both the spatial and temporal dimensions. Spatially, the interactions were analysed for the whole bus line and locally at the event venue. Temporally, the interactions were analysed on a day (6.00 to 23.00) with a one-hour time window and on a shorter focus interval with a 15-minutes time window. This approach of analysis at multiple granularity levels supports a correct understanding of patterns as demonstrated by Laube and Purves (2011). The analysis involved computing summary statistics on interactions aggregated by time intervals defined by the time window considered. Then, we used these statistics to identify outliers which point to the boundaries of the event lifespan.

Firstly, the *number of stoppings near the event venue* and the *balance between approaching and moving-away stoppings* were computed for each 1-hour interval. From the variation of these two variables the occurrence of an event, if any, has been detected and its start and end time estimated. Each estimated time was associated with an uncertainty time interval. Then, we computed the *number of stoppings near the event venue* for each time sub-interval defined by a shorter (15-minutes) time window within the two event start and end time intervals. By comparing the variation of current *number of stoppings near the event venue* and the corresponding normal values we refined the estimated event start and end times. The same comparison allowed revealing the temporal patterns of arrival and departure of event attendees.

The application of the analysis framework on the bus data and the two categories of events (large scale and medium scale) allowed answering the analysis questions we had formulated. We were able to detect both categories of events on the days they occurred (see Figure 5.6 and Figure 5.7 for a large scale event, Figure 5.10 and Figure 5.11 for a medium scale event). We were able to confirm the absence of events on days where they did not occur (17/11/2012 for a large scale event shown in Figure 5.4 and Figure 5.5, and 28/11/2012 for a medium scale event that was not shown). Also, we were able to estimate the start and end times of a detected event with some error margin of around 30 minutes and reveal the temporal patterns of arrival and departure of event attendees (see Figure 5.8 and Figure 5.12). However, the results from this experiment show that the success level of the framework in detecting the two categories of events differs. The bigger the event, the easier it is detected. For example, while the rugby match on 24/11/2012 in the Aviva stadium was detected with its start and end time well estimated (see Figure 5.7), the concert on 17/11/2012 in the National concert hall was detected and its start time estimated but the end time was not (see Figure 5.11).

The application of the proposed analysis framework in this experiment required a prior knowledge of the event location and the event to be of large scale. Regarding event detection, this case appears simplified compared to a general case aiming at detecting events of any size without prior knowledge of the location. Such general case has been commonly handled using social media data such as in (Sklar et al., 2012; Dashdorj et al., 2016). However, the objective in this experiment was a wider analysis of mobility pattern. Despite the requirement to have prior knowledge of event location, this wider analysis generates knowledge that is useful in the neighbourhood of the known location. For example, the knowledge about the temporal pattern of arrival and departure of event attendees can be useful in the planning and management of future events.

## 5.2 Experiment 2 – spatial dynamics of a dynamic geographic context

In this experiment we aim at demonstrating the applicability of the proposed analysis framework to cases in which the context element cannot be considered to be zero-dimensional. In such cases, spatial dynamics of the context need to be considered in addition to its temporal dynamics. This is the case of a geographic context of type "space" (e.g., a road network, a polygonal subdivision representing the temperature in different regions). Specifically, the experiment intends to discover some characteristics of the geographic context from the movement patterns. The case studied in the experiment determines the traffic congestion characteristics of a bus route over time by analysing a dynamic attribute of the bus movement. This section (5.2) is based on (Mazimpaka and Timpf, 2016*b*).

### 5.2.1 Datasets

In this experiment, a subset of movement data that have been used in experiment 1 and described in section 5.1.1 is analysed. The subset considered is made of trajectories of buses on bus line 4, which has been shown in Figure 5.1(a). We prepared the movement data through a series of transformation and cleaning operations. The operations include discarding GPS points located far away from the bus route, computing missing values, and identifying journeys. A journey is a sub-trajectory from one route end to the other. The final dataset contained 2,249 journeys made on 28 days between November 2012 and January 2013 including weekdays and weekends. Each journey was assigned a unique identifier and labelled with its direction. The route of bus line 4 represented by a sequence of road segments and bus stops along them is used as geographic context data for analysing the bus movement data.

### 5.2.2 Extraction of movement interactions

The main geographic context in this experiment is the road on which buses move. A road is one-dimensional and covers several point positions. Considering it as a dynamic context means that it has a dynamic attribute in which the analysis is interested. In this case, the dynamic attribute can have different values at different locations on the same road. Extracting interactions of the bus with the entire road would be very inaccurate because of these spatial dynamics. Therefore we started with a further pre-processing in which we partition the route into small segments on which local mobility patterns can be discovered.

Since bus stops are distributed along the route, they can be used to partition it. We partitioned the route by splitting the road segment between two successive bus stops as follows. Let `R` denote the bus route containing `n` bus stops, and `Mid(`$\widehat{st}$`)` denotes the midpoint of the curve between two points `s` and `t`.

$R =< p_1, p_2, \ldots, p_n >$ , where $\quad p_i$ is the $i_{th}$ bus stop

The $j_{th}$ route segment is represented as:

$$S_j = \begin{cases} \widehat{p_1 q} & , \text{ with } \quad q = Mid(\widehat{p_j p_{j+1}}) \quad \text{if} \quad j = 1 \\ \widehat{q p_n} & , \text{ with } \quad q = Mid(\widehat{p_{j-1} p_j}) \quad \text{if} \quad j = n \\ \widehat{q q'} & , \text{ with } \quad q = Mid(\widehat{p_{j-1} p_j}), \quad q' = Mid(\widehat{p_j p_{j+1}}) \quad \text{otherwise} \end{cases}$$

The operation of partitioning the route resulted in 65 road segments corresponding to the 65 bus stops of the route. After partitioning the route into small road segments, each labelled with the identifier of the bus stop on it, we extracted instances of the interactions of *passing* these road segments. To this end, the closest bus stop of each journey point has been used to assign the point a road segment. On each journey, the chronologically ordered list of points assigned a specific road segment constitute the interaction of *passing* it. Each *passing* interaction has the following attributes: start time, end time, identifier of the road segment being passed and the identifier of the journey (and implicitly identifier of the bus) passing the road segment. The start and end times of each *passing* interaction are the timestamps respectively of the first and last points of the sequence making it.

### 5.2.3 Analysis of movement interactions

The analysis of interactions involved the selection of a dynamic attribute of the movement and the analysis of its variation during interactions. We selected the *delay* attribute, which contains the number of seconds by which the bus is delayed at each position recorded. From the *delay* attribute we derive a new attribute *delayChange*, which is the difference between the delay value at the current position and its immediate predecessor. The *delayChange* attribute reflects the characteristics of a location and time period better than the *delay* attribute. The reason is that the change in delay is more closely related to the current location and time than the *delay*, which might have been accumulated from far previous locations and times. The delay change has one the following three meanings: the just travelled road segment negatively affected the movement (*delayChange>0*), or did not affect the movement (*delayChange=0*), or affected the movement positively (*delayChange<0*). In this experiment, the first case (i.e., delay increase) was considered potential to reflect important characteristics of road segments.

In order to explore the variation of the dynamic attribute *delayChange* during *passing* interactions, we aggregated its values on the route segments and within one hour time intervals. The aggregation was done by averaging available values. The aggregated values can be explored from different temporal granularity levels. For example, we proposed different views that can support a visual analysis of the aggregated values. These views show the values in a binary encoding (`delay increase` vs. `no delay increase`) to simplify the visual analysis while emphasising the importance attached to the case of delay increase. The *Hour-Location* view shows the variation in different hours and among the route segments on a selected day. The *Day-Location* view aggregates the *passing* interactions that occurred in one hour on each route segment for different days of the week for a selected one hour interval.

The *Hour-Location* view is used in Figure 5.13 to show the variation of delay change in different hours and among the route segments (represented by corresponding bus stop numbers) on the 13/12/2012. From the bus movement pattern shown in Figure 5.13 we can note that the segments represented by bus stops 13 and 31 were passed with an increase of delay for almost the whole day as indicated by the red dot on different hours of the day. Another observation is that the buses passed the route segments represented by the first bus stops (1-5) generally without increase of delay. The movement patterns shown in this view can be compared to the patterns shown in the *Day-Location* view. The *Day-Location* view in Figure 5.14 shows the variation of delay change among road segments and days of the week within three one-hour intervals (7:00 to 7:59 (a), 8:00 to 8:59 (b), and 9:00 to 9:59 (c)). It can be noted that the same route segments represented by bus stops 13 and 31 are passed with a delay increase within these three one-hour intervals on all the days of the week (shown on `Y` axis labelled with 0 to 6 for Sunday to Saturday).

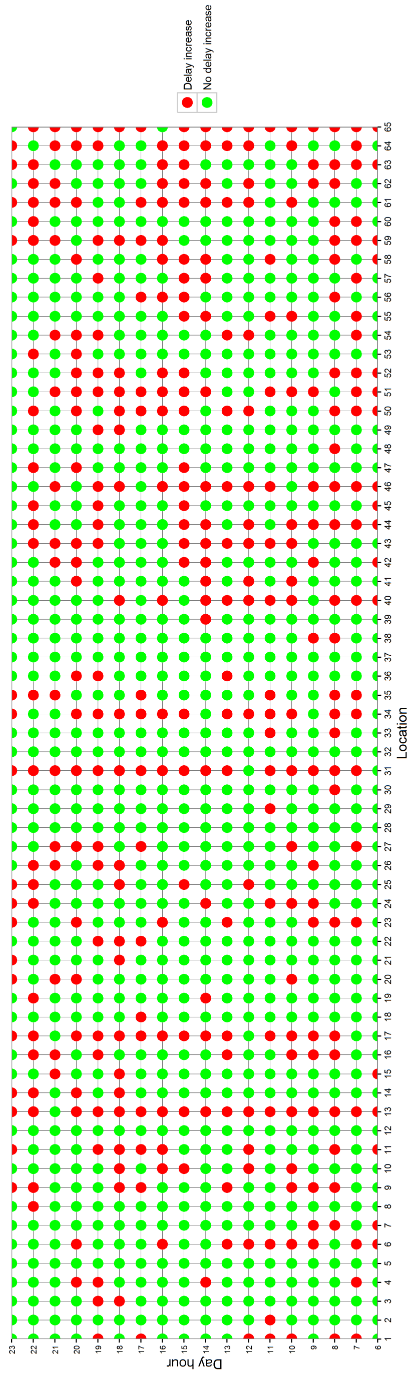
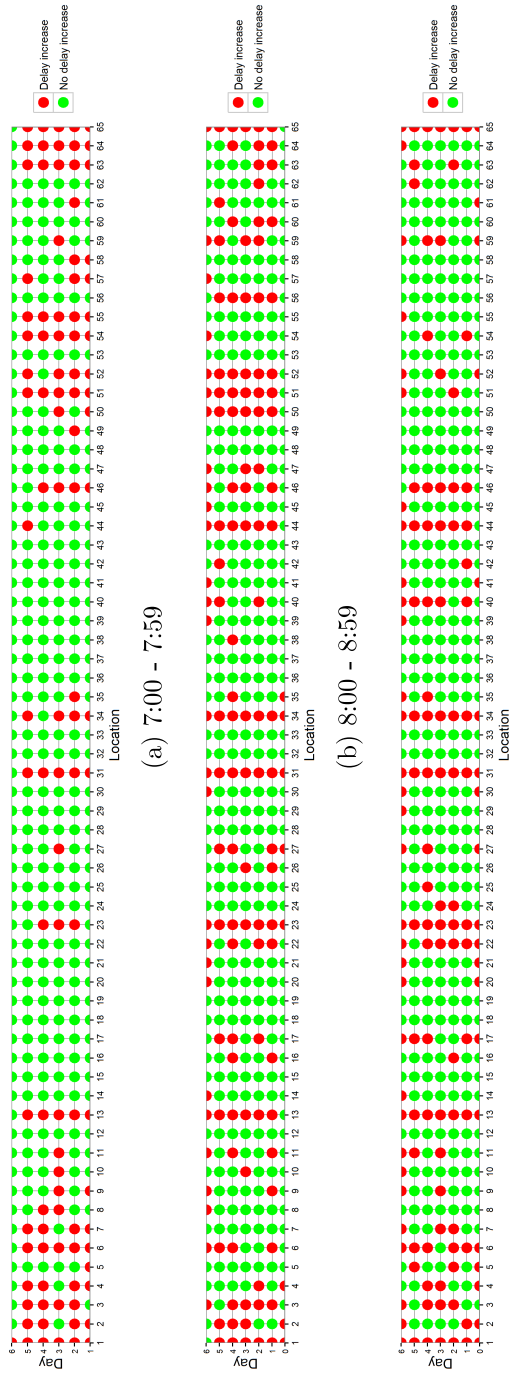



Fig. 5.13 Variation of delay change among one-hour intervals and route segments on 13/11/2012




(a) 7:00 - 7:59

(b) 8:00 - 8:59

(c) 9:00 - 9:59

Fig. 5.14 Variation of delay change among days of the week and route segments within selected one-hour intervals

We have discussed an approach for exploring the variation of a dynamic movement attribute during interactions of the bus and its route. However, the overall problem addressed in this analysis is to discover general characteristics of locations along the bus route based on the variation of a dynamic movement attribute. We turned this overall problem into a classification problem: we considered five classes of traffic congestion situation into which we had to classify the different route segments for the different one-hour intervals of the day. The solution approach follows the standard steps of classification: extraction of discriminative features and classification based on these features.

We put the days of the week into three categories exhibiting different mobility pattern (Weekday, Saturday, and Sunday) to which we associate three features; *wdweight, satweight*, and *sunweight* respectively. These features are weights of the congestion level on a specific day category. For each "one hour – one route segment" unit we compute each of the three features using the following equation:

$$dweight = \frac{n_d}{n} . \frac{(n-n_{min})}{n}, n \geq n_{min}$$

where `dweight` is the feature corresponding to a certain day category `k` (i.e., `dweight` is *wdweight* or *satweight* or *sunweight*), and $\mathtt{n_d}$ is the number of days of category `k` on which the unit was passed with a delay increase. In this equation, `n` is the total number of days of category `k` on which the unit was passed (including both with and without delay increase). The parameter $\mathtt{n_{min}}$ is a selected minimum number of days of category `k` the unit must have been passed to be considered for analysis. This parameter is used to ensure that we discover a general pattern. In case an analysis unit has been passed on a few days (e.g., one or two) all with a delay increase, the value of `dweight` would be the maximum while the passing frequency is not enough for decision. Therefore, $\mathtt{n_{min}}$ is used to avoid this bias. It ensures that if a unit has been passed on a few days (a parameter to be set), it is not considered in the analysis. The equation implies that the more the days a unit is passed with a delay increase, the higher the weight will be.

Based on the three features and a selected threshold value $\mathtt{w_{min}}$ we devised a set of eight ($2^3$) rules that we use in a rule-based classification for classifying each "one hour – one route segment" unit into one class of traffic congestion situation:

*IF wdweight* $> w_{min}$ *AND satweight* $> w_{min}$ *AND sunweight* $> w_{min}$ *THEN sclass='Always'*

*IF wdweight* $> w_{min}$ *AND satweight* $> w_{min}$ *AND sunweight* $\leq w_{min}$ *THEN sclass='Weekdays and Saturdays'*

*IF wdweight* $> w_{min}$ *AND satweight* $\leq w_{min}$ *AND sunweight* $> w_{min}$ *THEN sclass='Always'*

*IF wdweight* $> w_{min}$ *AND satweight* $\leq w_{min}$ *AND sunweight* $\leq w_{min}$ *THEN sclass='Weekdays only'*

*IF wdweight* $\leq w_{min}$ *AND satweight* $> w_{min}$ *AND sunweight* $> w_{min}$ *THEN sclass='Weekend only'*

*IF wdweight* $\leq w_{min}$ *AND satweight* $\leq w_{min}$ *AND sunweight* $> w_{min}$ *THEN sclass='Weekend only'*

*IF wdweight* $\leq w_{min}$ *AND satweight* $> w_{min}$ *AND sunweight* $\leq w_{min}$ *THEN sclass='Weekend only'*

*IF wdweight* $\leq w_{min}$ *AND satweight* $\leq w_{min}$ *AND sunweight* $\leq w_{min}$ *THEN sclass='Rarely/never'*

Each rule compares each of the three features with the threshold value to decide the class to which the unit is assigned. For example, the first rule means that a unit is classified as always congested if all weights (weekday, Saturday, and Sunday) are above the threshold value.

We extracted discriminative features (*wdweight, satweight*, and *sunweight*) on each "bus stop – one-hour" unit and performed a classification of the units. Considering that the value of each feature varies between 0 and 1, we took $\mathtt{w_{min}}$=0.5 as a threshold to separate high feature values (indicating high weights) from low ones. We also considered that our analysis unit should have been passed on more than two days to be considered for classification. Therefore, we took $\mathtt{n_{min}}$=2. The classification was done by evaluating the classification rules on the three features for each unit. The classification gave the final analysis result, which is a characterisation of the segments that make the bus route. We had pre-defined five classes of traffic congestion situation: `Always` (always congested), `Weekdays and Saturdays`, `Weekdays only`, `Weekend only`, and `Rarely/never` (rarely or never congested). Each segment of the route at each hour of the day is classified in one of these traffic congestion situations. Finally, the result of this classification can be visualised to facilitate understanding of the overall analysis result (e.g., see Figure 5.15) and possibly initiate a further analysis.

The *Hour-Location* view is used in Figure 5.15 to show a classification of different locations and time periods according to their traffic congestion situation. This classification represents discovered characteristics of the context element. For example, Figure 5.15 shows some route segments that are almost always congested (e.g., segments represented by bus stop 31 and 34). The figure further shows that in early morning hours (between 6:00 and 7:59) only few route segments are congested and this congestion happens on weekdays only. The location characteristics discovered and the interpretation given to movement patterns can be crosschecked by visualising the data of a selected time period in the *Map view*, which is a geographical view. For example, the time periods shown by the snapshots of Figure 5.16 conform to the finding that segments represented by bus stops 37 and 41 are congested between 9:00 and 10:00 in weekends. Furthermore, in the framework of supporting the crosschecking of results the *Map view* allows adding other available geographic data that were not initially considered as part of context data. For example, the snapshots shown in the *Map view* in Figure 5.16 show also the location of recreation facilities (the Aviva stadium in the current extent). Figure 5.16 shows that the stadium is located near the road segment represented by bus stop 42. After adding the location of recreation facilities we can note that most of the locations that are identified as congested only in weekends (see Figure 5.15) are near the Aviva stadium. This hints to weekend events in the stadium and its neighbourhood as the possible cause for traffic congestion in these locations.

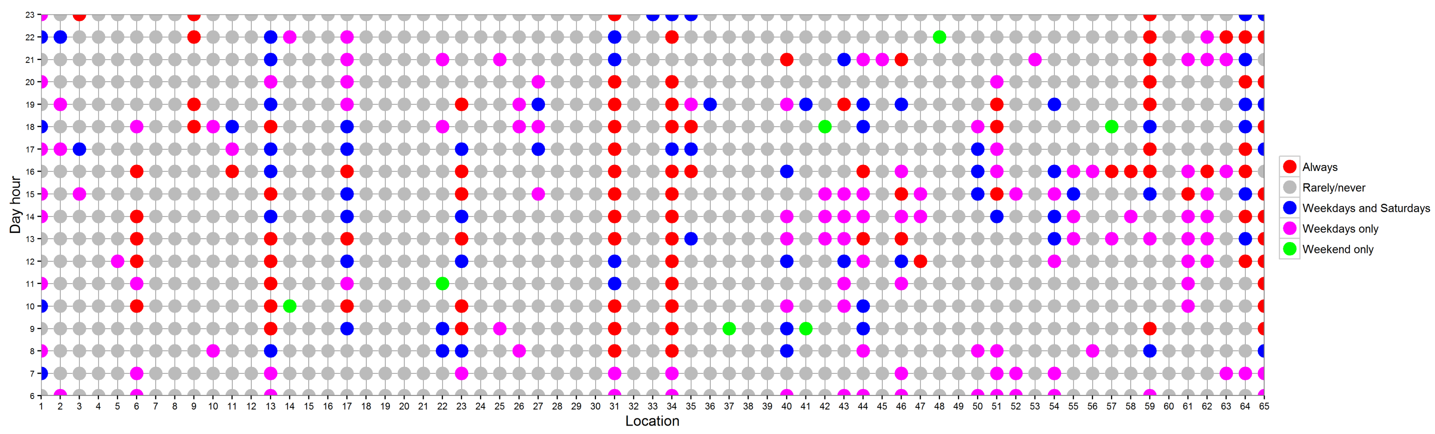



Fig. 5.15 Discovered traffic congestion situations of different locations along the bus route

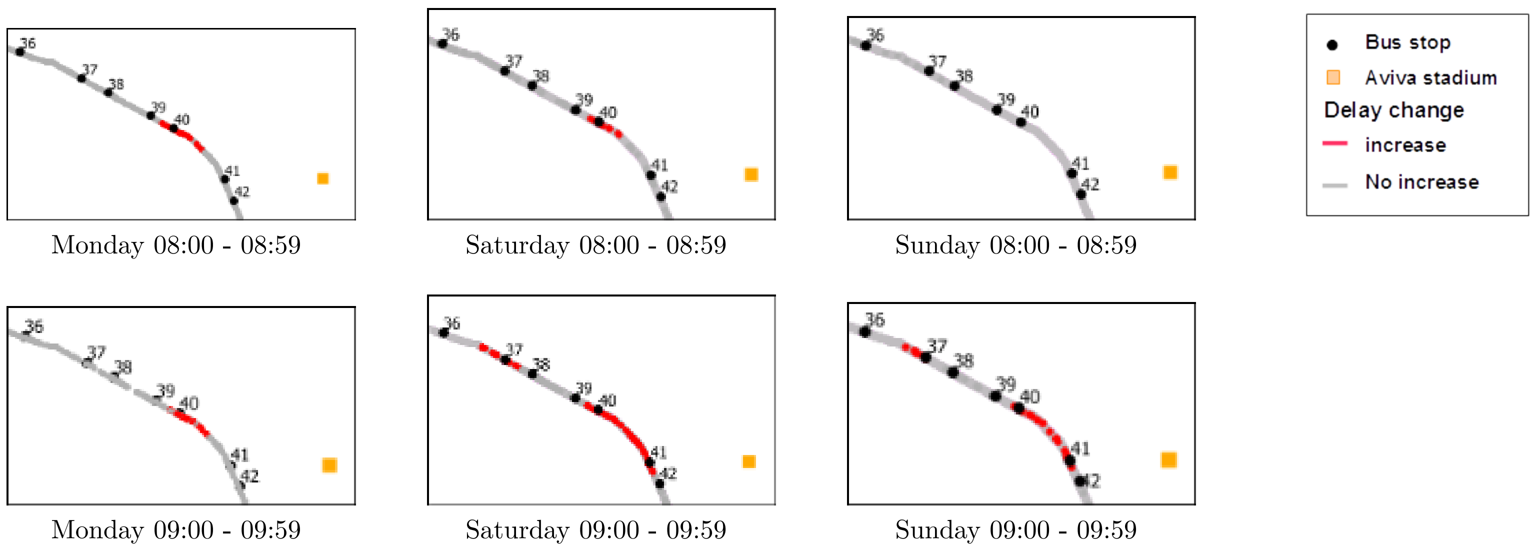



Fig. 5.16 Variation of delay change (increase vs. no increase) along some segments of the bus route within selected time periods (Map view)

### 5.2.4 Summary and discussion of results of experiment 2

This experiment has focused on exploring the spatial dynamics of a dynamic geographic context in addition to its inherent temporal dynamics. A study case involving a geographic context element of type "space" has been chosen. The reason behind the choice is that a context element of type "space" is a good example that presents spatial and temporal dynamics, both being equally important. Specifically, the study case aimed at analysing bus movement data taking into account the dynamics of the bus route. Unlike a big social event, which can be considered to have one point position as done in experiment 1 (section 5.1), a route (represented by a sequence of road segments) extends over several point positions that can have different characteristics.

We analysed the bus movement data by applying the analysis framework proposed in chapter 4. In order to take into account the spatial dynamics of the context we proposed partitioning the context element into small units for which local characteristics can be considered. Specifically, we partitioned the bus route into small segments such that each segment is allocated one bus stop which serves to identify it. Instead of extracting interactions with the entire route, we extracted interactions between the bus and the route segments. We identified the *passing* interaction (i.e., the bus passing a route segment) as suitable for the analysis case in this experiment. We extracted the *passing* interactions and attached to each interaction the start and end times, the identifier of segment being passed and the identifier of the journey (and hence the bus) passing it. In the next step, the analysis of extracted interactions was carried out by exploring the variation of a dynamic movement attribute during interactions. To this end, the values of the dynamic attribute "delay change" have been aggregated per route segment and per one-hour time interval. We have then extracted three features from the aggregated values and used them to perform a classification. The classification was intended to categorise each route segment at each hour of the day in one of five predefined classes of traffic congestion situations: `always congested`, `congested on weekdays and Saturdays`, `congested on weekdays only`, `congested on weekend only`, and `rarely or never congested`.

The analysis carried out in this experiment produces the two results expected from the proposed analysis framework (see Figure 4.2). These are an interpretation of the movement pattern observed and a description of some characteristics of the geographic context. The classes of traffic congestion situation to which each route segment is assigned at different hours of the day constitute characteristics of the context that the analysis discovers. The association of a movement pattern with route segments having a specific characteristic (revealed by the classification) is part of interpreting the movement pattern. For example, a slow movement in weekends along route segments found to be congested only on the weekend and located near a stadium can be observed. The association of this movement pattern with the locations having such characteristics allows interpreting the pattern as a traffic congestion caused by weekend events in the stadium and its neighbourhood. Likewise, in the experiment it has been found (from the classification) that route segments that are congested in early morning hours (6:00 to 7:59) are congested only on weekdays. This can lead to interpreting a slow movement on some route segments within this period as congestion mainly caused by the movement of people to workplaces.

The study case considered analysed movement data of buses with predefined route. While this case can be seen as simple compared to other cases involving context elements of type "space" the approach proposed is general to be applied to these other cases. For example, the analysis carried out in this experiment can be done on movement data of other vehicles without predefined route (e.g., taxis and private cars). In this case, a route can be defined as a sequence of road segments between a fixed origin and destination. Then, the route can be partitioned into segments of equal length or based on background knowledge, and a suitable dynamic movement attribute (e.g., speed) selected. In the next step, aggregation (spatial and temporal) can be performed on the movement data recorded on the selected route only. Although the *delay* attribute used in the study case was an existing attribute recorded along the movement, the dynamic attribute to be analysed can be a derived attribute. For example, the *delay* attribute could be derived from the position timestamp and the bus schedule. Likewise, the *speed* can be derived from the distance travelled between two successive positions and the time used.

The dynamic geographic context may cover an area instead of being a linear feature such as the road considered in this experiment. For example, we can think of a study case investigating the link between the mobility pattern of vehicles and the air pollution level in a city. In this case the context element can be modelled as a polygon representing an urban region and the level of air pollution in the region. It is this "air pollution level" attribute, which makes this context dynamic. A common approach for measuring air pollution over an area (e.g., a city) involves installing monitoring stations which take measurements to be considered for the surrounding area (Sharker and Karimi, 2014). For the whole city, a simple approach would take the average of the values at different stations. However, as explained earlier in this experiment, the wider or the longer the context element is the more inaccurate will be the result of analysis considering it as one unit. Therefore, our approach based on partitioning the context element into small units is also applicable to this case of an urban area. The urban area can be partitioned into small regions (represented as polygons) based on the distribution of air pollution measurement stations. Then, the *passing* interactions of vehicles with these small regions can be extracted and analysed. During the analysis, the values of the pollution level (a dynamic attribute of the context) are taken from local measurements. For the analysis of interactions, the speed with which the small regions are passed can be considered for a dynamic movement attribute.

## 5.3 Experiment 3 – Using context data for post-processing of movement patterns

The preceding experiments have used context data at the pattern discovery step of the movement data analysis. In the direction of answering the research question about the analysis step at which context data should be integrated, we use in this experiment context data at a different step for comparison. The context data are used for supporting the interpretation of already discovered movement patterns. Furthermore, while the preceding experiments used as context data information from event listing webpages and public geographic data, we use geo-social media data in this experiment for comparison purposes. It is important to note that, unlike the preceding experiments, this experiment deals with a static context. The focus is not on the dynamics of the context but on experimenting with a different form of context data and a different analysis step at which context data are integrated into the analysis process. This section (5.3) is based on (Mazimpaka and Timpf, 2015).

### 5.3.1 Datasets

In this experiment the movement data of taxi cabs in San Francisco are analysed. The dataset was downloaded from the CRAWDAD website (Piorkowski et al., 2009). It includes trajectories of 536 taxi cabs recorded in 22 days, from 18 May to 8 June 2008. For each point position in the trajectories, the recorded data are geographic coordinates (latitude and longitude), timestamp, and the taxi occupancy status. The data were recorded at an average sampling rate of 60 seconds. The whole dataset contained around 11 million GPS points, which were pre-processed through outlier removal, coordinate system transformation, and restriction to San Francisco Bay area. Since only the GPS points corresponding to a change of taxi occupancy status (pick-up and drop-off) are useful for the case study, this subset was selected for further processing.

The subset contained 432,618 pick-up points and 425,996 drop-off points. Furthermore, considering that the mobility pattern is generally repeated every week, a one-week subset was selected for analysis in order to avoid unnecessary heavy computation. In order to keep the mobility coverage of the whole study area, the final subset included the data of all taxis in operation during the time period considered.

The context data are georeferenced Flickr photos taken in the same area and same time period as the movement data. Each photo is associated with descriptive information (title, description, and tags) which carry the semantics needed to support movement data analysis. The application case intends to discover and explain taxi frequent stop patterns. That is, to detect if in the study area there are regions where taxi cabs often stop, to locate these regions and explain the reason of frequent stops. We argue that *frequent stop* patterns can be discovered and located from the trajectories alone but that there are different reasons of frequently stopping in different regions. The reason of frequently stopping in a specific region can only be found by integrating context data of the study area into the analysis process.

### 5.3.2 Pattern discovery

The frequent stop patterns of taxi cabs are detected using a clustering method, which groups closely located taxi pickup and drop-off points irrespective of their timestamp. A density based clustering method, namely DBSCAN (Ester et al., 1996), has been selected for the following reasons. The number of regions where stopping frequently occurs is not initially known and these regions generally have irregular shapes. DBSCAN meets these requirements because it does not require the number of clusters to be specified as input and it can discover irregularly shaped clusters.

The selected sub-dataset was stored in PostgreSQL/PostGIS DBMS and clustered using the DBSCAN algorithm implemented in the Weka data mining toolkit. For clustering parameters, we selected 0.02 and 5 for the neighbourhood distance (`Eps`) and the minimum number of points (`MinPts`) respectively. The results of this clustering are shown in Figure 5.17. Since the density of points in the downtown area is very high, the points in this area produced one very big cluster which needs to be disaggregated to identify different small clusters hidden in it. The very big cluster was disaggregated by a second level clustering with different parameter values (0.015 and 10 for `Eps` and `MinPts` respectively). The values of clustering parameters were selected based on the exploration of the dataset and after experimenting with different values. The final clusters were transformed into regions by creating concave hulls enclosing cluster points. These regions (see Figure 5.18) represent the *frequent* *stop* patterns that were discovered. In order to make these patterns useful for some application (e.g. updating land use maps, decision support for infrastructure development ...), they need to be interpreted. The interpretation is done by adding to each region the information that can explain why taxi cabs frequently stop in it. This calls for integrating geographic context data, which is done in the next step; pattern analysis.

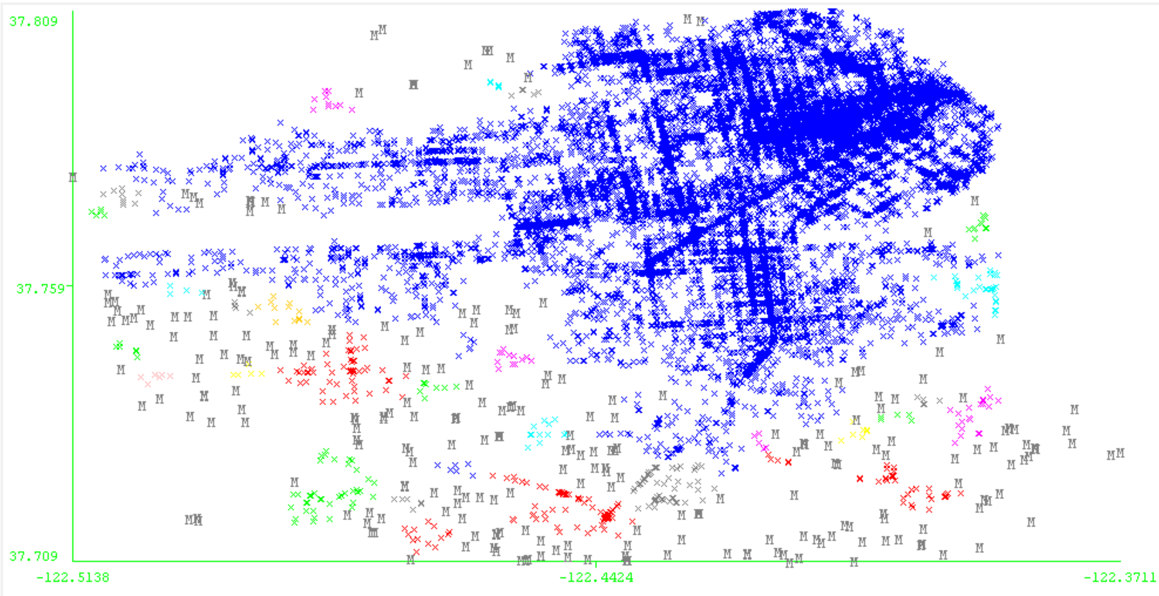



Fig. 5.17 Intermediate clustering results

### 5.3.3 Pattern analysis

In order to interpret the frequent stop patterns discovered in the preceding step we acquired and used geographic context data in the extent of the regions representing the patterns. In this experiment we used context data from a social media platform with the aim of exploring the feasibility of its use as a source of geographic context data for movement analysis. Among the available social media platform we selected Flickr for the following reasons. The San Francisco Bay area is among the best represented geographic regions in Flickr as it has been reported in previous work (Rattenbury et al., 2007). Among Geo-social media platforms, Flickr presents the fewest restrictions to data access (Spinsanti et al., 2013). We wrote a PHP script that uses Flickr API to download descriptive information associated with photos in a specified geographic area.

We defined some categories of common functions of regions in urban area (`Educational, Commercial, Residential, Recreational`, and `Institutional`). Then, we selected

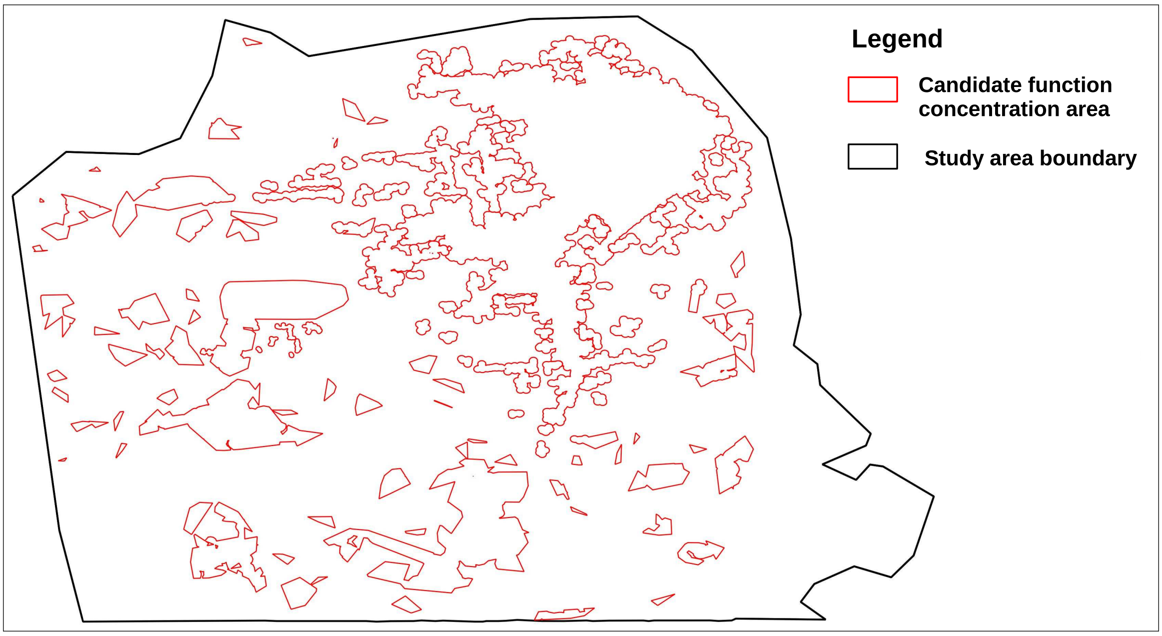



Fig. 5.18 Regions representing the frequent stop patterns discovered

in each category some common concepts based on which photos were downloaded. For example, under the category "Educational" photos with tags such as `university`, `school`, `kindergarten`, `classroom`, and `faculty` were downloaded in all regions. For each photo $P_i$ retrieved under the specification of functional category $F_j$, we computed a variable we called *functional weight* as follows:

$$w_{P_i} = d/e$$

where $d$ is the number of tags on $P_i$ which are concepts under functional category $F_j$ and $e$ is the total number of tags on photo $P_i$.

We defined the *functional vector* of a region as a vector made of normalised value of different function categories in it. For each region $R_k$, the entry in the functional vector corresponding to the functional category $F_l$ is computed as follows:

$$v_{kl} = \frac{\sum_{i=1}^{n} w_i}{\sum_{j=1}^{m} w_j}$$

where $n$ is the number of photos in region $R_k$ which are associated with the functional category $F_l$ and $m$ is the total number of photos of all functional categories in $R_k$ while $w_i$ (or $w_j$) is the weight functional of respective photo $P_i$ (or $P_j$).

Finally, we classified each region in a functional category based on the values in its functional vector. In this experiment the region was assigned the category with the highest value in the functional vector. Alternatively, a fuzzy classification could be performed. In this case, the region is assigned a ranked list of the first two or three functional categories with highest value. The classification of regions produces *function concentration areas*; i.e., areas where specific urban functions are concentrated (see Figure 5.19). These classified regions provide the needed interpretation of the *frequent stop* patterns that were discovered. For example, a region labelled as a residential area means that frequent stopping in it is for bringing people home or taking them from home to other places.

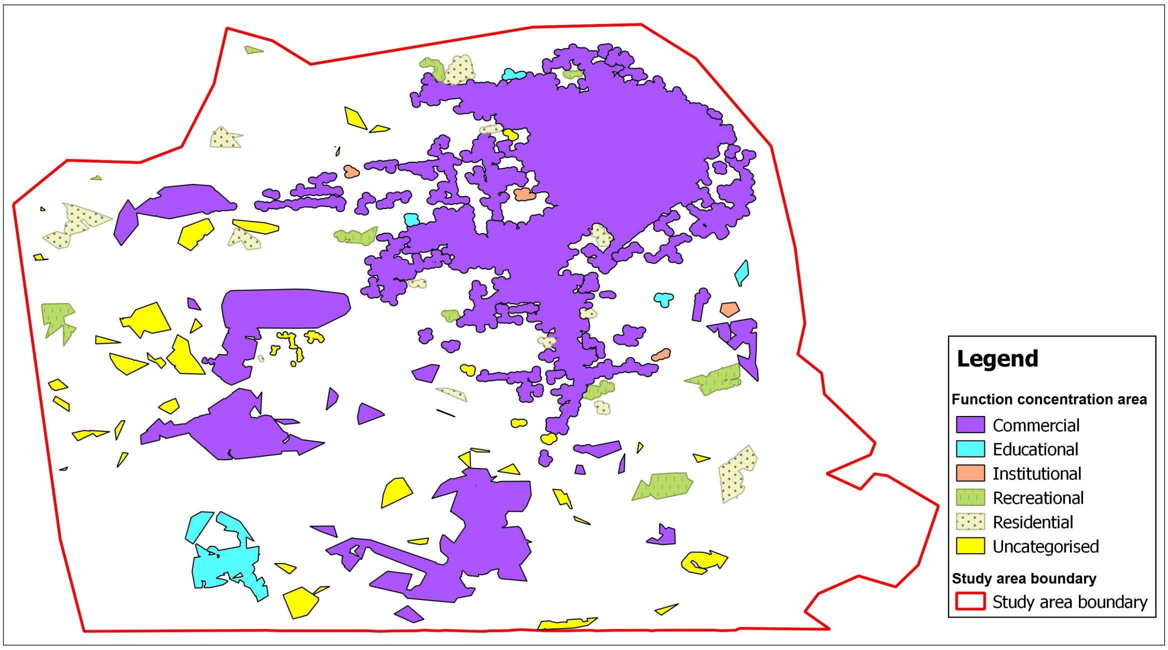



Fig. 5.19 Categorised regions as an interpretation of the discovered patterns

## 5.3.4 Summary and discussion of results of experiment 3

In this experiment, the potential of integrating geographic context data at the pattern interpretation level of movement data analysis has been explored. Unlike previous experiments, this experiment discovered movement patterns without involving geographic context data and then used geographic context data for post-processing the discovered patterns. Specifically, the case study of the experiment was intended to discover frequent stop patterns of taxi cabs (i.e., locations where taxi cabs recurrently stop) and to interpret them. The *frequent stop* patterns were discovered using a data mining method, namely clustering. For interpreting the *frequent stop* patterns, geographic context data from a geo-social media platform, namely Flickr were used. Geo-Social media contain descriptive information potential for interpreting movement patterns. However, their distribution in regions is skewed with some areas having no data while some others have a lot of data. The skewed distribution of social media has many different causes including the skewed technology penetration and the different level of interest in different regions. For example, Flickr photos are largely taken in places with known attractions. On the other hand, they are rarely taken in residential areas and most of those taken in residential areas are not shared publicly. This explains why movement patterns discovered in some regions may not get matching social media and hence cannot be interpreted using social media from a selected geo-social media platform. For example, the regions classified as `uncategorised` in Figure 5.19 could not be interpreted using Flickr photos.

# Chapter 6

# Discussion

The different experiments presented in chapter 5 addressed different aspects of integrating the geographic context into movement data analysis. These aspects include different types of context elements, different models for these context elements, different types of sources of geographic context data, and different analysis steps at which the context data are integrated into the analysis. These aspects reflect the research questions addressed in this thesis. In this chapter, I revisit the research questions and discuss how they were addressed by reflecting on the aspects dealt with in the experiments and the overall framework proposed in chapters 3 and 4. The chapter ends with a discussion of the limitations of the work done.

## 6.1 Results in relation to research questions

### 6.1.1 Modelling a dynamic geographic context and relating it to movement

The context for movement is very broad such that a study considering it needs to set the scope. For this thesis, the scope has been set to dynamic geographic context. The first task in the thesis was about clarifying, representing and relating this context to movement as expressed in the first research question:

> *RQ 1: How can we model a dynamic geographic context so as to allow relating it to the movement it embeds?*

Even the dynamic geographic context presents varieties. With this consideration, I started by categorising the elements that make the geographic context and identifying the properties that can make it a dynamic one. Four categories were distinguished:

*static objects, moving objects, events* and *space*. Considering that an element in each of these categories is a geographic phenomenon, I adopted the standard geographic data models (vector-based and field-based) for modelling the geographic context. The application of these different models on the four categories provides a general modelling approach, which include common cases and cases that are not common but possible. For example, while a moving object modelled as a point is common, a moving polygon and a moving line have been rarely studied. The modelling approach proposed in this thesis covers also these uncommon cases as context for movement with examples such as an oil spill and sea waves. The context elements in the proposed categories except the *space* category are commonly represented using vector-based models. The modelling approach proposed in this thesis is more general as it considers also the field-based model. This allows accommodating cases where the context data are acquired in the form of images. The modelling approach proposed in chapter 3 is general and therefore, the choice of the model to use depends on the size and shape of the element, and the scale considered by the application.

The modelling approach proposed for a dynamic geographic context does not reflect the dynamic nature of the context as it considers a snapshot; i.e., it represents the context at a specific time instant. Therefore, this dynamic nature had to be considered in the approach of relating the context to the movement it embeds. To this end, I adopted a qualitative approach that uses the change of spatial relations over time, termed as *movement interaction*, to relate the context to movement. I proposed and formalised a small set of interactions depending on the type of context element involved. The very basic interactions proposed are observed between a moving object and a zero-dimensional (point) context element with a fixed location. These are: *approaching, arriving, stopping, leaving* and *moving-away*. By considering a higher dimension (line and area) and the change of location for the context element new interactions appear or existing ones change the meaning and hence the naming. The extension to the basic interactions contains the following interactions: *passing, encounter, meeting, separating, jointly-moving*, and *separately-moving*.

### 6.1.2 Sources of geographic context data and their integration into analysis

Movement data and geographic context data are generally acquired as different datasets. One of the objectives of this thesis was to identify sources of geographic context data and the analysis step at which they should be integrated. This objective was expressed in the following research question:

*RQ 2: What are the sources of geographic context data for movement analysis and at which analysis step should these data be integrated?*

In order to answer the first part of this question, I performed a literature review on different geo-referenced data and used different types of data identified in the experiments presented in chapter 5. The main forms of geographic context data that I identified based on their sources are presented in section 2.4.4. They are *other georeferenced datasets of the same area as the movement data, geo-social data, data from other sensors, trajectories of other moving objects*, and *other unstructured or semi-structured data*. In order to assess the feasibility of using these geographic context data to support movement analysis, I used them in the experiments presented in chapter 5. I used roads and bus stops as "other georeferenced data of the same region as movement data", and events data as "other unstructured data" in section 5.1. Data about different static objects have been used as "geo-social data" in section 5.3.

The literature review and the experiments showed that the sources identified contain specific types of geographic context elements and not the others. Georeferenced datasets of the region generally contain context data about static objects and space. For example, roads, points of interests, administrative subdivisions and environmental conditions of regions are available from this source. Social media contain a wide range of geographic context data mainly because several social media platforms exist with different foci. For example, social media platforms that focus on locations and their descriptions (e.g., Foursquare and Flickr) are a common source of data about static objects (e.g., POIs). On the other hand, social media platforms that focus on sharing latest information (e.g., Twitter and Facebook) are a common source of data about events. Other unstructured or semi-structured data commonly available on webpages are also a source of data about events and POIs. Data from sensors other than trajectory recorders constitute also a wide range of context data because several sensors recording different things can be used. For example, while the smartphone records the trajectory using the GPS sensor, it may record also the temperature (as environmental conditions) using a thermometer, and the presence of other moving objects using the Bluetooth sensor. As the name suggests, trajectories of other moving objects are context data about moving objects.

Another observation from the experiments is that human effort, needed in preparing the context data, varies among the identified sources. Structured context data (other georeferenced data of the region, trajectories of other moving objects, and geo-social media) require less background knowledge compared to unstructured data from webpages and the analyst's mind. As a result, the integration of the former can be more easily automated than the integration of the latter.

In order to identify the step at which the geographic context data should be integrated into the analysis, I executed experiments with context data integration at different steps of the KDD process so as to compare the results. The steps of the KDD process were grouped into three main phases: data preparation, data mining or pattern discovery, and pattern interpretation and evaluation. Since the two datasets come from different sources they are prepared differently. Therefore, the data preparation phase is not concerned by the context data integration. The experiments in sections 5.1 and 5.2 integrated context data at the pattern discovery step while the experiment in section 5.3 used context data at the pattern interpretation and evaluation phase. When context data are integrated at pattern discovery phase, only patterns that can be interpreted in the subsequent analysis step are discovered because each pattern embeds context information. On the other hand, when patterns are discovered first and then context data integrated for interpreting them some of the discovered patterns may find no match in the available context data. In that case, they remain non-interpreted despite the processing effort that has been put in their discovery. For example, in the experiment presented in section 5.3 some regions of *frequent stop* patterns were discovered but could not be interpreted (see uncategorised regions in Figure 5.19).

In this thesis, context data have been considered essential for discovering meaningful patterns and therefore the proposed framework integrates them at the pattern discovery step. However, it is acknowledged that in some cases context data can be integrated at the pattern interpretation phase; especially in case the movement data contain some additional attributes that provide semantics useful in pattern discovery. For example, the additional attributes can support the filtering of input to focus on data that are likely to contain the sought pattern such as done in experiment 3 with the "occupancy status" attribute (see section 5.3). This approach of using geographic context data for post-processing discovered patterns has been generally used for a geographic context which is static (e.g., the POIs in (Furletti et al., 2013) and (Siła-Nowicka et al., 2016)).

Different geographic phenomena located in the proximity of a moving object constitute the geographic context for its movement. The decision on which geographic context element to consider depends on the analysis aimed at. The major objective of movement analysis integrating geographic context data is to explore the relation (e.g., influence) between the movement and the context element. With this objective, the analyst generally knows on which movement and which context element his interest is. Based on this knowledge, suitable context data are acquired or the only available context data decide the type of analysis and the context element involved. In an exploratory study without predefined context element of interest, based on background knowledge the analyst may consider the context element most likely to influence the movement. For example, a big social event (e.g., a concert) is likely to influence the movement of a public transportation vehicle while a small social event (e.g., a group meeting) is not. Again, the context element is considered only if related data can be obtained.

### 6.1.3 Contextualised pattern discovery and analysis

The overall aim in relating movement to its embedding dynamic geographic context is to support interpretation of movement patterns and further understanding of the context. After proposing a conceptual model for relating movement to its embedding dynamic geographic context, the next task was to use this model to reach the overall aim. This task was expressed in the following two research questions:

*RQ 3: How can the relation between the movement and its embedding dynamic geographic context be explored in space and time to support movement pattern interpretation?*

*RQ 4: How can the relation between the movement and its embedding dynamic geographic context be explored in space and time to support a deeper understanding of the geographic context?*

In order to address these research questions, I proposed an analysis framework based on the KDD process for movement data and related dynamic geographic context data. This framework is presented in chapter 4 and evaluated in the experiments presented in chapter 5. The framework comprises two phases, namely *extraction of movement interactions* and *analysis of movement interactions*, which together correspond to the last two phases of the KDD process (data mining, and pattern interpretation/evaluation).

In the phase of interaction extraction, spatial analysis and data mining methods are used to compute interactions as movement patterns embedding context information. For example, in the experiment presented in section 5.1, instead of simple *stop* patterns, *stopping* interactions are extracted. They represent *stop* patterns at specific context elements (e.g., a bus stop, or an event venue). The *stopping* interactions have been extracted using a density-based clustering and a computation of spatial proximity. Each interaction extracted is associated with attributes including the identifier of the context element involved in the interaction and zero or more dynamic attributes. For example, the *passing* interaction extracted in section 5.2 is associated with an identifier of the road segment being passed and a dynamic attribute "delay change" indicating the change of delay incurred while passing the road segment.

In the phase of *interaction analysis*, statistical and/or data mining methods are used to quantify the extracted interactions and explore the variation of any associated dynamic attributes. The thesis proposed different methods for carrying out this quantification and exploration. The methods proposed include aggregation per space and time, computation of summary statistics, detection of outliers, and sequence analysis. For example, in section 5.1.3 the variation of summary statistics of interactions has been compared to established bounds to identify temporal outliers.

The proposed analysis framework leads to one or two results. The first result is a high level description of movement patterns that allows answering questions such as why the objects move the way they do or how the objects move given some physical constraints. By producing this result, the proposed framework is an answer to the third research question (RQ 3). For example, the unusually frequent stopping of the buses at bus stops in one direction and not the other in a specific period (see the experiment presented in section 5.1) was interpreted to be caused by the movement of event attendees. The second possible result from the framework is new knowledge about the geographic context. This result makes the framework an answer to the fourth research question (RQ 4). For example, the exploration of the variation of the dynamic attribute "delay change" associated with the interaction of "passing" a road segment (see the experiment presented in section 5.2) found that this variation correlates with the traffic congestion level of the road segment. This analysis resulted in characterising each road segment by identifying periods in which it is generally congested. This characteristic of the road segment constitutes new knowledge generated about the geographic context element (road segment) because the input context data were only the geometry of the road segment and the identifier of the bus stop on it.

The thesis proposed a general analysis framework that uses the conceptual model presented in chapter 3. However, the implementation of the framework may raise issues specific to the case being handled, which is defined by the type of context element involved and the model selected for it. Important issues and how the thesis addressed them are discussed next:

The dynamic geographic context element of type “space” may be too long or too wide such that aggregate functions on interactions involving it hide local patterns. For example, averaging different speed values recorded while passing a road may result in an average speed which is very different from the actual speed at many locations along the road. The thesis proposed alleviating this problem by subdividing the long or wide geographic context element into small units on which aggregation functions can be applied. This approach has been applied in the experiment presented in section 5.2 for identifying locations of recurring traffic congestion and corresponding congestion time. The approach is applicable to both the case of a linear “space” (e.g., experiment in section 5.2) and an areal “space”. An example of the latter case can be investigating the link between the mobility pattern of vehicles and the air pollution level in a city. A way of applying the proposed approach to this case has been briefly discussed in the last paragraph of section 5.2.4.

The dynamic geographic context element of type “event” exists for a time interval shorter than the time period covered by the study. This leads to the challenge that at some time instants the movement cannot be related to it because it does not exist. The main objective in relating the movement to an event is to understand the effect of the event on the mobility pattern. The objective can be also to learn characteristics of the event from the mobility pattern (e.g., given the origin of event attendees, determine the type of event (Calabrese et al., 2010)). These objectives generally require studying and comparing the mobility pattern “before”, “during”, and “after” the event, in a way similar to how Bagrow et al. (2011) studied the effect of events on phone call activity. For the “during” phase the movement is related to the event, while for the phases “before” and “after” it is related to the event location. Some analysis cases involve first discovering the event and proceeding to studying its effect or characteristics. In this case, the event location can be discovered from suitable context data such as news from webpages (as done in (Bagrow et al., 2011)) and social media (as done in (Sklar et al., 2012)). Once the event location and a rough estimation of its time (e.g., the day) are known, the movement can be related to the event location and the relation explored in time intervals covering a long period (e.g., a whole day). The experiment presented in section 5.1 exemplifies and demonstrates this case where the event has to be discovered first.

## 6.2 Limitations and issues remaining to be addressed

The thesis has answered the initially defined research questions as presented in section 6.1. However, some issues remain. These issues are discussed next.

### 6.2.1 Completeness of the interaction set

The set of movement interactions proposed in this thesis is not complete for representing more complex changes of spatial relations between the movement and a dynamic geographic context. The thesis focused on a small set of interactions that can be unambiguously defined and differentiated. However, there are other complex changes of spatial relations that cannot be expressed using the interactions defined in the thesis. For example, the change in which the moving object bypasses the context element cannot be expressed using the interactions defined.

The common characteristic to the interactions defined so far in this thesis is that the moving object moves in 1D. This does not mean that it follows a straight line. It means that at any time there are only two directions in which the object can move. As real cases, it is assumed that a vehicle follows the road that passes closest to the event venue. Likewise, it is assumed that the predator is aware of the location of the prey and follows the shortest path that can lead to catching it. The proposed conceptual model covers all the possible interactions in such a 1D-based movement. In order to cover the remaining interactions there is a need to increase the number of degrees of freedom of movement such that at any time the moving object can move in any direction. This extension will cover cases where, at a road crossing, the vehicle can take any possible road. It will cover also cases where an animal can move freely with respect to another (hence covering for example the diverging movement).

### 6.2.2 Consideration of all types of dynamics of a dynamic geographic context

Different aspects that can make a geographic context dynamic have been discussed in section 3.1. In the conceptual framework proposed and the experiments carried out, the dynamics due to change of location of the context element, change of attribute value, and bounded lifespan have been considered. However, the dynamics due to change of the extent of the context element (its size and/or shape) have not been considered. As an example of such case, one can think of the movement of a ship and an oil spill expanding on the sea around its epicentre. Similarly, one can think of the movement of a car and a street demonstration which expands as more people join or shrinks as some people leave. The change of size and/or shape of the context element add another challenge to relating the context to movement. Furthermore, the change of the extent of the context element may be associated with the other changes that have been discussed and considered in this thesis. One of the issues to be addressed is to choose the suitable temporal window for aggregation operations such as those proposed in the analysis framework (see section 4.3). The uncertainty in the representation of the boundary of the context element also needs to be considered. As discussed by Von Groote-Bidlingmaier and Timpf (2012), the change of extent of moving areal features is influenced by the elements or characteristics of the local geography (e.g., wind direction for an oil spill). Therefore, the consideration of further geographic context elements can support the choice of parameters for handling uncertainties (e.g., the size of tolerance area for the boundary of the context element).

### 6.2.3 Consideration of all models of the dynamic geographic context

The conceptual framework proposed in this thesis considers both vector-based and field-based models for the context. However, the experimental evaluation has been carried out on vector-based models of the context only. An evaluation on the context modelled as a field is missing. This evaluation would uncover issues specific to this model and propose necessary adjustments to the implementation of the general framework for addressing them.

# Chapter 7

# Conclusions and outlook

This chapter summarises the contribution of the thesis and gives the direction for future work. In section 7.1, the main contribution of the thesis is summarised. Section 7.2 revisits the research hypothesis with reference to the main contribution while section 7.3 discusses how the work will be extended in the future.

## 7.1 Main contribution

With the increased capabilities for tracking moving entities, the analysis of movement data has become a hot research topic. The discovery of movement patterns hidden in large volumes of movement data is one of the major themes that have attracted attention. While the initial work considered movement data in a pure geometric form, the trend is now on integrating context data to make the discovered pattern more relevant to application fields. It is in this direction that this thesis aimed to contribute. Specifically, the thesis aimed at proposing a comprehensive approach for integrating a dynamic geographic context into the analysis of movement data.

The analysis methods proposed in this thesis can be usefully applied in different fields where the movement environment needs consideration, especially because of its dynamics. For example, location based services (e.g., traffic information and routing services) can benefit from predictions and updates on local traffic conditions. By integrating changes occurring on road segments and their neighbourhood, these methods can enable such predictions and updates.

With the initial aim of the thesis, its main contribution can be summarized in the following points.

### A conceptual framework for modelling a dynamic geographic context and relating it to movement

This thesis has clarified *a dynamic geographic context* and delineated its meaning from the general concept of *context*. A context which has a geographic location and a time at which it exists or existed is geographic. When the geographic context is associated with changes over time, it becomes a dynamic geographic context. The thesis discussed the different types of changes that can be associated with a dynamic geographic context. Considering the varieties that exist in things that can be considered as a dynamic geographic context, the thesis established a *typology of dynamic geographic context elements*. Then, based on the standard models for geographic data, the thesis proposed an approach for modelling the different types of geographic context elements. The thesis proposed a conceptual model of *movement interactions* to account for the dynamics of context elements. The concept of movement interactions abstracts the change of spatial relations between the moving entity and the context element. A set of basic movement interactions has been proposed and these interactions have been organized in *conceptual neighbourhood graphs*. The conceptual neighbourhood graphs can support qualitative reasoning during movement data analysis.

### A methodological framework for exploiting the relation between the movement and the context to support: a) the interpretation of movement patterns, and b) a deeper understanding of the context

The relation between the movement and the context implies that they can affect each other. This thesis has proposed a methodological framework for exploring this relation to identify and describe the mutual effect between the movement and the context. The framework, based on the knowledge discovery process (KDD), comprises two main phases: *interaction extraction* and *interaction analysis*. The thesis proposed different methods that can be used in each of these phases and some *analysis operations* that can be carried out on movement interactions. The output from the analysis framework comprises two results. The first result is a high level description of the observed movement patterns which constitute the interpretation of the pattern. The second result is a description of the movement context which constitutes new knowledge about it and hence provides a deeper understanding of the movement context.

In addition to proposing an analysis framework, the thesis made two further related contributions. Firstly, the thesis *surveyed the sources of geographic context data for movement analysis*. Context data from some of the sources identified have been used in experimental evaluation of the framework. From the experimental evaluation, challenges associated with some sources have been identified and some approaches for addressing them discussed. Secondly, the thesis *studied the suitability of different analysis phases as a step for integrating context data.* The phases assessed are data mining, at which context data are intended to support pattern discovery, and pattern interpretation/evaluation, at which they are intended to support the post-processing of patterns already discovered. The suitability study was done by executing experiments in which the context data were integrated at the respective analysis phases. The results showed that in some cases patterns can be discovered first and then context data used for post-processing them. Such cases are common for stop patterns (also called movement suspension patterns (Orellana and Wachowicz, 2011)) which can be easily discovered from movement parameters (Dodge et al., 2008). However, in such case some of the patterns discovered may find no match in the available context data which will make their interpretation difficult or impossible. Furthermore, some patterns can be discovered only if context data are considered. For example, group patterns (see section 2.3.2) cannot be discovered from the trajectory of one moving entity; trajectories of other moving entities have to be considered. Therefore, both the pattern discovery and the pattern interpretation/evaluation are potential steps for integrating geographic context data into movement analysis. The choice of the integration step depends on the data being analysed and the type of pattern being mined.

## 7.2 Revisiting the hypothesis

The work carried out in this thesis tested the following hypothesis which was stated at the beginning of the research:

> *Identifying and analysing interactions between a moving object and the dynamic geographic context supports understanding of the movement patterns and the movement context.*

Based on the results achieved which have been discussed in chapter 6, I can confirm the statement of the hypothesis. The conceptual framework presented in chapter 3 formalises interactions between a moving object and the dynamic geographic context. The analysis framework discussed in chapter 4 extracts and analyses these interactions. The evaluation of the applicability of the framework, as presented in chapter 5, demonstrated that the framework correctly produces the two expected results (see Figure 4.2).

The first experiment (see section 5.1) extracted and analysed *stopping, approaching*, and *moving-away* interactions between buses and context elements (bus stop points, and a big social event). The analysis discovered, for example, a bus movement pattern characterised by increased frequency of stopping at some bus stop points and normal frequency at others within a specific time period. The consideration of a known big event location allowed interpreting this movement pattern as the movement of event attendees. At the same time, the context which was initially a potential big event (represented by the event location) was later confirmed to be an event occurrence. The detected event was further described by discovering its start and end times and the pattern of arrival and departure of its attendees (i.e., progressive, on time, delay ...). This demonstrates the possibility of supporting understanding of movement patterns through the resulting interpretation. It demonstrates also the possibility of supporting understanding of the movement context through the new knowledge generated about it (e.g., the start and end time of the event, the pattern of arrival and departure of its attendees).

The second experiment (see section 5.2) extracted and analysed the *passing* interaction of buses with route segments. The analysis assigned each route segment a class of traffic congestion situation for each one-hour interval (e.g., between 9:00 and 10:00 the route segment identified by 37 is “congested only in weekends”). At the same time the analysis discovered a bus movement pattern characterised by a slow motion always in weekends. The consideration of the route segments on which these locations are found showed that they are segments congested always on the weekend and located near a stadium. This allowed interpreting the slow movement pattern as a traffic congestion caused by events in the stadium or its neighbourhood. As in the first experiment, this interpretation supports understanding the movement pattern. The congestion situation class assigned to each route segment constitutes knew knowledge about the context (route segment) which allows further understanding of the context.

## 7.3 Outlook

On the basis of the limitations and issues remaining to be addressed as discussed in section 6.2, the following points are proposed for future work.

### Extending the set of movement interactions

In order to model and analyse more complex changes of relations between movement and the dynamic geographic context, the set of basic interactions proposed in this thesis needs to be extended. The extension will enable representing and modelling complex change of relations such as a moving entity bypassing a context element, and a moving entity converging with another moving entity. The interactions in the proposed conceptual model have been defined based on changes of distance relations only. A direction to be explored for extending this set of interactions is to consider other spatial relations (e.g., directional relations) in addition to distance relations. Directional relations have been combined with topological relations in (Salamat and Zahzah, 2012) to model motion events between two moving objects. Some of the motion events modelled in (Salamat and Zahzah, 2012) are similar to the interactions defined in this thesis between a moving object and a moving context element. Therefore, future work can explore how the notion of F-histograms (Matsakis and Wendling, 1999), used in (Salamat and Zahzah, 2012), can be applied and generalised on different types of context elements.

#### Evaluating the proposed analysis framework on the context modelled as a field

The different standard models for geographic phenomena have been proposed for modelling the geographic context. These are the object-based and field-based models. However, all the experiments executed to evaluate the analysis framework have used geographic context data in object-based models. Nevertheless, some context elements such as temperature and wind speed are better modelled as a field. Although a conversion from field-based to vector-based representation is possible, future work should evaluate the framework on a geographic context modelled as a field. In this case the best way of measuring distances needs to be identified; for example between the Euclidean distance and the Manhattan distance (Deza and Deza, 2009). The change of attribute value is a common cause of the dynamics of a field-based context. While exploring the variation of the attribute within time intervals as proposed in the analysis framework, a missing value problem may be observed. This problem can be addressed using interpolation methods such as those used in (Mitas et al., 1997; Dragicevic and Marceau, 2000).

#### Extending the conceptual model to consider the dynamics due to change of the extent of a context element

The thesis has identified different factors that cause the dynamics of a dynamic geographic context. These are; the change of location, a bounded lifespan, the change of value of a thematic attribute, and the change of extent. However, the dynamics due to changing extent have not been considered in the thesis. Future work should study the dynamics due to changing extent (e.g., changing size and changing shape) both alone and combined with the other dynamics that have been considered in the thesis. One of the challenges to be addressed in this case is the uncertainty of the boundary of the context element. For addressing this challenge, the concept of a *spatial range* used in (Praing and Schneider, 2007) and the uncertainty threshold used in (Mokhtar and Su, 2004) are an important direction to explore.

### Supporting the integration of geographic context data

Different sources of context data have been identified in this thesis. It is common that a specific source may not contain context data for a specific region (e.g., Flickr has no photos for some regions as shown in section 5.3). It would be interesting if a tool can be developed to support the integration of context data from multiple sources. If for a specific region no context data are available from a specific source, the tool can switch to a different source to find data with same properties for the same region. The tool could further prepare the data to fit the format used. For example the tool developed for downloading data from Flickr (see section 5.3 ) can be extended to download data from different social media platforms (e.g., Flickr and Foursquare).

# References


Andersson, M., Gudmundsson, J., Laube, P. and Wolle, T. (2008), 'Reporting leaders and followers among trajectories of moving point objects', *GeoInformatica* **12**(4), 497–528.
**URL:** *http://dx.doi.org/10.1007/s10707-007-0037-9*

Andrienko, G., Andrienko, N. and Heurich, M. (2011), 'An event-based conceptual model for context-aware movement analysis', *International Journal of Geographical Information Science* **25**(9), 1347–1370.
**URL:** *http://dx.doi.org/10.1080/13658816.2011.556120*

Andrienko, G., Andrienko, N., Hurter, C., Rinzivillo, S. and Wrobel, S. (2011), From movement tracks through events to places: Extracting and characterizing significant places from mobility data, *in* 'Proceedings of IEEE Conference on Visual Analytics Science and Technology (Providence, RI, USA, 23 - 28 October 2011)', IEEE, pp. 161–170.

Andrienko, N. and Andrienko, G. (2013), 'Visual analytics of movement: An overview of methods, tools and procedures', *Information Visualization* **12**(1), 3–24.
**URL:** *http://dx.doi.org/10.1177/1473871612457601*

Andrienko, N., Andrienko, G., Wachowicz, M. and Orellana, D. (2008), Uncovering interactions between moving objects, *in* T. J. Cova, H. J. Miller, K. Beard, A. U. Frank and M. F. Goodchild, eds, 'Extended Abstracts of GIScience 2008', Springer, Berlin-Heidelberg, pp. 16–26.

Ankerst, M., Breunig, M. M., Kriegel, H.-P. and Sander, J. (1999), Optics: Ordering points to identify the clustering structure, *in* 'Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data', SIGMOD '99, ACM, New York, NY, USA, pp. 49–60.
**URL:** *http://doi.acm.org/10.1145/304182.304187*

Asher, N. and Sablayrolles, P. (1994), 'A compositional spatio-temporal semantics for french motion verbs and spatial PPs', *Semantics and Linguistic Theory* **4**(0), 1–15.
**URL:** *http://journals.linguisticsociety.org/proceedings/index.php/SALT/article/view/2453*

Augustine, D. J. and Derner, J. D. (2013), 'Assessing herbivore foraging behavior with GPS collars in a semiarid grassland', *Sensors* **13**(3), 3711–3723.
**URL:** *http://www.mdpi.com/1424-8220/13/3/3711*

Baglioni, M., Fernandes de Macêdo, J., Renso, C., Trasarti, R. and Wachowicz, M. (2009), Towards semantic interpretation of movement behavior, *in* M. Sester, L. Bernard and V. Paelke, eds, 'Advances in GIScience: Proceedings of the 12th AGILE Conference', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 271–288.

Bagrow, J. P., Wang, D. and Barabási, A.-L. (2011), 'Collective response of human populations to large-scale emergencies', *PLOS ONE* **6**(3), 1–8.
**URL:** *https://doi.org/10.1371/journal.pone.0017680*

Bandeira de Melo, L. F., Lima Sábato, M. A., Vaz Magni, E. M., Young, R. J. and Coelho, C. M. (2007), 'Secret lives of maned wolves (chrysocyon brachyurus illiger 1815): as revealed by GPS tracking collars', *Journal of Zoology* **271**(1), 27–36.
**URL:** *http://dx.doi.org/10.1111/j.1469-7998.2006.00176.x*

Benkert, M., Gudmundsson, J., Hübner, F. and Wolle, T. (2008), 'Reporting flock patterns', *Computational Geometry* **41**(3), 111 – 125.
**URL:** *http://www.sciencedirect.com/science/article/pii/S092577210700106X*

Birant, D. and Kut, A. (2007), 'ST-DBSCAN: An algorithm for clustering spatial–temporal data', *Data &amp Knowledge Engineering* **60**(1), 208 – 221.
**URL:** *http://www.sciencedirect.com/science/article/pii/S0169023X06000218*

Bleisch, S., Duckham, M., Galton, A., Laube, P. and Lyon, J. (2014), 'Mining candidate causal relationships in movement patterns', *International Journal of Geographical Information Science* **28**(2), 363–382.
**URL:** *http://dx.doi.org/10.1080/13658816.2013.841167*

Bolbol, A., Cheng, T., Tsapakis, I. and Haworth, J. (2012), 'Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification', *Computers, Environment and Urban Systems* **36**(6), 526 – 537.
**URL:** *http://www.sciencedirect.com/science/article/pii/S0198971512000543*

Bolchini, C., Orsi, G., Quintarelli, E. and Tanca, L. (2011), 'Context modeling and context awareness: steps forward in the context-ADDICT project', *IEEE Data Engineering Bulletin* **34**(2), 47–54.

Brakatsoulas, S., Pfoser, D., Salas, R. and Wenk, C. (2005), On map-matching vehicle tracking data, *in* 'Proceedings of the 31st International Conference on Very Large Data Bases', VLDB '05, VLDB Endowment, pp. 853–864.
**URL:** *http://dl.acm.org/citation.cfm?id=1083592.1083691*

Buchin, M., Dodge, S. and Speckmann, B. (2014), 'Similarity of trajectories taking into account geographic context', *Journal of Spatial Information Science* **2014**(9), 101–124.

Buchin, M., Kruckenberg, H. and Kölzsch, A. (2013), Segmenting trajectories by movement states, *in* S. Timpf and P. Laube, eds, 'Advances in Spatial Data Handling: Geospatial Dynamics, Geosimulation and Exploratory Visualization', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 15–25.

Cagnacci, F., Focardi, S., Heurich, M., Stache, A., Hewison, A., Morellet, N., Kjellander, P., Linnell, J. D., Mysterud, A., Neteler, M. et al. (2011), 'Partial migration in roe deer: migratory and resident tactics are end points of a behavioural gradient determined by ecological factors', *Oikos* **120**(12), 1790–1802.

Calabrese, F., Pereira, F. C., Di Lorenzo, G., Liu, L. and Ratti, C. (2010), The geography of taste: Analyzing cell-phone mobility and social events, *in* P. Floréen, A. Krüger and M. Spasojevic, eds, 'Pervasive Computing', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 22–37.

Cao, H., Mamoulis, N. and Cheung, D. W. (2005), Mining frequent spatio-temporal sequential patterns, *in* 'Proceedings of the Fifth IEEE International Conference on Data Mining', ICDM '05, IEEE Computer Society, Washington, DC, USA, pp. 82–89.
**URL:** *http://dx.doi.org/10.1109/ICDM.2005.95*

Cao, H., Mamoulis, N. and Cheung, D. W. (2007), 'Discovery of periodic patterns in spatiotemporal sequences', *IEEE Transactions on Knowledge & Data Engineering* **19**(4), 453–467.
**URL:** *http://dx.doi.org/10.1109/TKDE.2007.1002*

Chon, J. and Cha, H. (2011), 'Lifemap: A smartphone-based context provider for location-based services', *IEEE Pervasive Computing* **10**(2), 58–67.

Clementini, E., Felice, P. D. and Oosterom, P. v. (1993), A small set of formal topological relationships suitable for end-user interaction, *in* 'Proceedings of the Third International Symposium on Advances in Spatial Databases', SSD '93, Springer-Verlag, London, UK, UK, pp. 277–295.
**URL:** *http://dl.acm.org/citation.cfm?id=647223.718898*

Cohn, A. G., Bennett, B., Gooday, J. and Gotts, N. M. (1997), 'Qualitative spatial representation and reasoning with the region connection calculus', *GeoInformatica* **1**(3), 275–316.
**URL:** *http://dx.doi.org/10.1023/A:1009712514511*

Cohn, A. G. and Hazarika, S. M. (2001), 'Qualitative spatial representation and reasoning: An overview', *Fundamenta informaticae* **46**(1-2), 1–29.

Dashdorj, Z., Tsogtbaatar, B., Tumurchudur, A. and Altangerel, E. (2016), High level event identification in social media, *in* 'Proceedings of the 12th International Conference on Semantics, Knowledge and Grids (SKG)', IEEE, pp. 121–125.

de By, R. (2011), *Principles of geographic information systems*, 2 edn, ITC, Enschede, The Netherlands.

de Lucca Siqueira, F. and Bogorny, V. (2011), 'Discovering chasing behavior in moving object trajectories', *Transactions in GIS* **15**(5), 667–688.
**URL:** *http://dx.doi.org/10.1111/j.1467-9671.2011.01285.x*

Demšar, U. and Virrantaus, K. (2010), 'Space–time density of trajectories: exploring spatio-temporal patterns in movement data', *International Journal of Geographical Information Science* **24**(10), 1527–1542.
**URL:** *http://dx.doi.org/10.1080/13658816.2010.511223*

Demšar, U., Buchin, K., van Loon, E. E. and Shamoun-Baranes, J. (2015), 'Stacked space-time densities: a geovisualisation approach to explore dynamics of space use over time', *GeoInformatica* **19**(1), 85–115.
**URL:** *http://dx.doi.org/10.1007/s10707-014-0207-5*

Dewulf, B., Neutens, T., Vanlommel, M., Logghe, S., Maeyer, P. D., Witlox, F., Weerdt, Y. D. and de Weghe, N. V. (2015), 'Examining commuting patterns using floating car data and circular statistics: Exploring the use of new methods and visualizations to study travel times', *Journal of Transport Geography* **48**, 41 – 51.
**URL:** *http://www.sciencedirect.com/science/article/pii/S0966692315001398*

Dey, A. K. (2001), 'Understanding and using context', *Personal Ubiquitous Computing* **5**(1), 4–7.
**URL:** *http://dx.doi.org/10.1007/s007790170019*

Deza, M. M. and Deza, E. (2009), *Encyclopedia of distances*, Springer.

Ding, L., Yang, J. and Meng, L. (2015), Visual analytics for understanding traffic flows of transport hubs from movement data, *in* 'Proceedings of the 27th International Cartographic Conference (Rio de Janeiro, Brazil, August 23-28, 2015)', pp. 23–28.

Dobson, E. D. and Bollinger, J. (1994), *Understanding Bollinger Bands*, Traders Press, Cedar Falls, Iowa, USA.

Dodge, S., Bohrer, G., Bildstein, K., Davidson, S. C., Weinzierl, R., Bechard, M. J., Barber, D., Kays, R., Brandes, D., Han, J. et al. (2014), 'Environmental drivers of variability in the movement ecology of turkey vultures (cathartes aura) in north and south america', *Philosophical Transactions of the Royal Society B* **369**(1643).

Dodge, S., Bohrer, G., Weinzierl, R., Davidson, S. C., Kays, R., Douglas, D., Cruz, S., Han, J., Brandes, D. and Wikelski, M. (2013), 'The environmental-data automated track annotation (env-data) system: linking animal tracks with environmental data', *Movement Ecology* **1**(1), 3.

Dodge, S., Weibel, R. and Lautenschütz, A.-K. (2008), 'Towards a taxonomy of movement patterns', *Information Visualization* **7**(3), 240–252.
**URL:** *http://dx.doi.org/10.1057/palgrave.ivs.9500182*

Dragicevic, S. and Marceau, D. J. (2000), 'A fuzzy set approach for modelling time in GIS', *International Journal of Geographical Information Science* **14**(3), 225–245.
**URL:** *http://dx.doi.org/10.1080/136588100240822*

Egenhofer, M. and Al-Taha, K. (1992), Reasoning about gradual changes of topological relationships, *in* A. U. Frank, I. Campari and U. Formentini, eds, 'Theories and methods of spatio-temporal reasoning in geographic space', Springer, Berlin Heidelberg, pp. 196–219.

Egenhofer, M. J. (1991), Reasoning about binary topological relations, *in* O. Günther and H.-J. Schek, eds, 'Advances in Spatial Databases: Proceedings of the 2nd Symposium on spatial databases (Zurich, Switzerland, August 28–30, 1991)', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 141–160.

Egenhofer, M. J., Clementini, E. and Di Felice, P. (1994), 'Topological relations between regions with holes', *International Journal of Geographical Information Science* **8**(2), 129–142.

Egenhofer, M. J. and Mark, D. M. (1995), 'Modelling conceptual neighbourhoods of topological line-region relations', *International journal of geographical information systems* **9**(5), 555–565.

Eriksen, A., Wabakken, P., Zimmermann, B., Andreassen, H. P., Arnemo, J. M., Gundersen, H., Milner, J. M., Liberg, O., Linnell, J., Pedersen, H. C., Sand, H., Solberg, E. J. and Storaas, T. (2009), 'Encounter frequencies between GPS-collared wolves (Canis lupus) and moose (Alces alces) in a scandinavian wolf territory', *Ecological Research* **24**(3), 547–557.
**URL:** *http://dx.doi.org/10.1007/s11284-008-0525-x*

Ester, M., Kriegel, H.-P., Sander, J. and Xu, X. (1996), A density-based algorithm for discovering clusters in large spatial databases with noise, *in* 'Proceedings of the Second International Conference on Knowledge Discovery and Data Mining', KDD'96, AAAI Press, pp. 226–231.
**URL:** *http://dl.acm.org/citation.cfm?id=3001460.3001507*

Fayyad, U. M., Piatetsky-Shapiro, G. and Smyth, P. (1996), From data mining to knowledge discovery: an overview, *in* U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, eds, 'Advances in Knowledge Discovery and Data Mining', American Association for Artificial Intelligence, Menlo Park, CA, USA, pp. 1–34.
**URL:** *http://dl.acm.org/citation.cfm?id=257938.257942*

Frank, A. U. (1991), Qualitative spatial reasoning with cardinal directions, *in* H. Kaindl, ed., 'Proceedings of the Seventh Austrian Conference on Artificial Intelligence (Wien, Austria, 24–27 September 1991)', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 157–167.

Frank, A. U. (1992), 'Qualitative spatial reasoning about distances and directions in geographic space', *Journal of Visual Languages & Computing* **3**(4), 343 – 371.
**URL:** *http://www.sciencedirect.com/science/article/pii/1045926X92900079*

Freksa, C. (1991), Conceptual neighborhood and its role in temporal and spatial reasoning, *in* M. Singh and M. L. Travé, eds, 'Decision Support Systems and Qualitative Reasoning: Proceedings of the IMACS International Workshop on Decision Support Systems and Qualitative Reasoning (Toulouse, France, 1991)', North-Holland, pp. 181–187.

Furletti, B., Cintia, P., Renso, C. and Spinsanti, L. (2013), Inferring human activities from gps tracks, *in* 'Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing', UrbComp '13, ACM, New York, NY, USA, pp. 5:1–5:8.
**URL:** *http://doi.acm.org/10.1145/2505821.2505830*

Gaffney, S. J., Robertson, A. W., Smyth, P., Camargo, S. J. and Ghil, M. (2007), 'Probabilistic clustering of extratropical cyclones using regression mixture models', *Climate Dynamics* **29**(4), 423–440.
**URL:** *http://dx.doi.org/10.1007/s00382-007-0235-z*

Galton, A. (1994), Lines of sight, *in* M. Keane, P. Cunningham, M. Brady and R. Byrne, eds, 'AI and Cognitive Science '94', Dublin University Press, pp. 103–113.

Galton, A. (2012), States, processes and events, and the ontology of causal relations., *in* M. Donnelly and G. Guizzardi, eds, 'Frontiers in Artificial Intelligence and Applications (FOIS)', Vol. 239, IOS Press, pp. 279–292.

Giannotti, F., Nanni, M., Pinelli, F. and Pedreschi, D. (2007), Trajectory pattern mining, *in* 'Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', ACM, New York, NY, USA, pp. 330–339.
**URL:** *http://doi.acm.org/10.1145/1281192.1281230*

Goker, A. and Myrhaug, H. I. (2002), User context and personalisation, *in* 'Proceedings of the 6th European Conference on Case Based Reasoning (Aberdeen, Scotland, 04 - 07 September 2002)', City University of London, pp. 4–7.

Gschwend, C. (2015), Relating movement to geographic context – Effects of preprocessing, relation methods and scale, phdthesis, University of Zurich.

Guo, H., Wang, Z., Yu, B., Zhao, H. and Yuan, X. (2011), Tripvista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection, *in* 'Proceedings of the 2011 IEEE Pacific Visualization Symposium (Hong Kong, China, 1-4 March 2011)', IEEE, pp. 163–170.

Han, J. and Miller, H. (2009), Geographic data mining and knowledge discovery: An overview, *in* H. Miller and J. Han, eds, 'Geographic Data Mining and Knowledge Discovery, Second Edition', CRC Press, pp. 1–26.
**URL:** *http://dx.doi.org/10.1201/9781420073980.ch1*

Hariri, A., Tabary, D., Lepreux, S. and Kolski, C. (2008), Context aware business adaptation toward user interface adaptation, *in* 'Communications of SIWN', Springer Verlag, pp. 46–52.

Helbing, D., Molnár, P., Farkas, I. J. and Bolay, K. (2001), 'Self-organizing pedestrian movement', *Environment and Planning B: Planning and Design* **28**(3), 361–383.
**URL:** *http://dx.doi.org/10.1068/b2697*

Jahnke, M., Ding, L., Karja, K. and Wang, S. (2017), Identifying origin/destination hotspots in floating car data for visual analysis of traveling behavior, *in* G. Gartner and H. Huang, eds, 'Progress in Location-Based Services 2016', Springer International Publishing, Cham, pp. 253–269.

Jeung, H., Yiu, M. L., Zhou, X., Jensen, C. S. and Shen, H. T. (2008), 'Discovery of convoys in trajectory databases', *Proceedings of the VLDB Endowment* **1**(1), 1068–1080.
**URL:** *http://dx.doi.org/10.14778/1453856.1453971*

Kalnis, P., Mamoulis, N. and Bakiras, S. (2005), On discovering moving clusters in spatio-temporal data, *in* C. Bauzer Medeiros, M. J. Egenhofer and E. Bertino, eds, 'Advances in Spatial and Temporal Databases: Proceedings of the 9th International Symposium on Spatial Temporal Databases (Angra dos Reis, Brazil, August 22-24, 2005)', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 364–381.

Kamber, M., Han, J. and Pei, J. (2012), *Data mining: Concepts and techniques*, 3rd edn, Morgan Kaufmann, Waltham, MA, USA.

Kang, J.-Y. and Yong, H.-S. (2010), 'Mining spatio-temporal patterns in trajectory data', *Journal of Information Processing Systems* **6**(4), 521–536.
**URL:** *http://dx.doi.org/10.3745/JIPS.2010.6.4.521*

Kays, R., Crofoot, M. C., Jetz, W. and Wikelski, M. (2015), 'Terrestrial animal tracking as an eye on life and planet', *Science* **348**(6240).
**URL:** *http://science.sciencemag.org/content/348/6240/aaa2478*

Keler, A. and Krisp, J. M. (2016), Visual analysis of floating taxi data based on interconnected and timestamped area selections, *in* G. Gartner, M. Jobst and H. Huang, eds, 'Progress in Cartography: EuroCarto 2015', Springer International Publishing, Cham, pp. 115–131.

Kraak, M. J. (2003), The space-time cube revisited from a geovisualization perspective, *in* 'Proceedings of the 21st International Cartographic Conference (Durban, South Africa, 10 - 16 August 2003)', Vol. 1995, The International Cartographic Association (ICA), pp. 1988–1996.
**URL:** *http://www.itc.nl/library/Papers_2003/art_proc/kraak.pdf*

Krisp, J. M. and Peters, S. (2011), 'Directed kernel density estimation (DKDE) for time series visualization', *Annals of GIS* **17**(3), 155–162.
**URL:** *http://dx.doi.org/10.1080/19475683.2011.602218*

Krisp, J. M., Peters, S. and Burkert, F. (2013), Visualizing crowd movement patterns using a directed kernel density estimation, *in* J. M. Krisp, L. Meng, R. Pail and U. Stilla, eds, 'Earth Observation of Global Changes (EOGC)', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 255–268.

Kurata, Y. and Egenhofer, M. J. (2006), The head-body-tail intersection for spatial relations between directed line segments, *in* M. Raubal, H. J. Miller, A. U. Frank and M. F. Goodchild, eds, 'GIScience 2006: Proceedings of the 4th International Conference on Geographic Information Science (Münster, Germany, September 20-23, 2006)', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 269–286.

Laube, P. (2009), Progress in movement pattern analysis, *in* B. Gottfried and H. Aghajan, eds, 'Behaviour Monitoring and Interpretation – BMI', Vol. 3, IOS Press, pp. 43–71.

Laube, P., Imfeld, S. and Weibel, R. (2005), 'Discovering relative motion patterns in groups of moving point objects', *International Journal of Geographical Information Science* **19**(6), 639–668.
**URL:** *http://dx.doi.org/10.1080/13658810500105572*

Laube, P. and Purves, R. S. (2011), 'How fast is a cow? cross-scale analysis of movement data', *Transactions in GIS* **15**(3), 401–418.
**URL:** *http://dx.doi.org/10.1111/j.1467-9671.2011.01256.x*

Laube, P., van Kreveld, M. and Imfeld, S. (2005), Finding REMO – detecting relative motion patterns in geospatial lifelines., *in* P. F. Fisher, ed., 'Developments in Spatial Data Handling', Springer, Berlin-Heidelberg, pp. 201–215.

Lee, J.-G., Han, J. and Whang, K.-Y. (2007), Trajectory clustering: A partition-and-group framework, *in* 'Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data', ACM Press, Beijing, China, pp. 593–604.
**URL:** *http://dx.doi.org/10.1145/1247480.1247546*

Li, Q., Zheng, Y., Xie, X., Chen, Y., Liu, W. and Ma, W.-Y. (2008), Mining user similarity based on location history, *in* 'GIS '08: Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems', ACM Press, Irvine, California, USA, pp. 1–10.
**URL:** *http://dx.doi.org/10.1145/1463434.1463477*

Li, Z., Ding, B., Han, J. and Kays, R. (2010), 'Swarm: mining relaxed temporal moving object clusters', *Proceedings of the VLDB Endowment* **3**(1-2), 723–734.
**URL:** *http://dx.doi.org/10.14778/1920841.1920934*

Li, Z., Ding, B., Han, J., Kays, R. and Nye, P. (2010), Mining periodic behaviors for moving objects, *in* 'KDD '10: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining', ACM Press, Washington, DC, pp. 1099–1108.
**URL:** *http://dx.doi.org/10.1145/1835804.1835942*

Long, J. A. (2015), 'Quantifying spatial-temporal interactions from wildlife tracking data: Issues of space, time, and statistical significance', *Procedia Environmental Sciences* **26**, 3 – 10.
**URL:** *http://www.sciencedirect.com/science/article/pii/S1878029615001711*

Long, J. A. and Nelson, T. A. (2013), 'Measuring dynamic interaction in movement data', *Transactions in GIS* **17**(1), 62–77.
**URL:** *http://dx.doi.org/10.1111/j.1467-9671.2012.01353.x*

Long, J. A., Nelson, T. A., Webb, S. L. and Gee, K. L. (2014), 'A critical examination of indices of dynamic interaction for wildlife telemetry studies', *Journal of Animal Ecology* **83**(5), 1216–1233.
**URL:** *http://dx.doi.org/10.1111/1365-2656.12198*

Lundblad, P., Eurenius, O. and Heldring, T. (2009), Interactive visualization of weather and ship data, *in* 'Proceedings of the 13th International Conference on Information Visualisation (Barcelona, Spain, 15-17 July 2009)', IEEE, pp. 379–386.

Maimon, O. and Rokach, L. (2010), Introduction to knowledge discovery and data mining, *in* O. Maimon and L. Rokach, eds, 'Data Mining and Knowledge Discovery Handbook', 2nd edn, Springer, New York, pp. 1–15.

Mamoulis, N., Cao, H., Kollios, G., Hadjieleftheriou, M., Tao, Y. and Cheung, D. W. (2004), Mining, indexing, and querying historical spatiotemporal data, *in* 'KDD '04: Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, Seattle, WA, USA, pp. 236–245.
**URL:** *http://dx.doi.org/10.1145/1014052.1014080*

Matsakis, P. and Wendling, L. (1999), 'A new way to represent the relative position between areal objects', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21**(7), 634–643.

Mazimpaka, J. D. and Timpf, S. (2015), Exploring the potential of combining taxi GPS and flickr data for discovering functional regions, *in* F. Bacao, M. Y. Santos and M. Painho, eds, 'AGILE 2015: Geographic Information Science as an Enabler of Smarter Cities and Communities', Springer International Publishing, Cham, pp. 3–18.

Mazimpaka, J. D. and Timpf, S. (2016*a*), 'Trajectory data mining: A review of methods and applications', *Journal of Spatial Information Science* **2016**(13), 61–99.

Mazimpaka, J. D. and Timpf, S. (2016*b*), A visual and computational analysis approach for exploring significant locations and time periods along a bus route, *in* 'Proceedings of the 9th ACM SIGSPATIAL International Workshop on Computational Transportation Science (Burlingame, California, October 31 - November 03, 2016)', ACM, New York, NY, USA, pp. 43–48.
**URL:** *http://doi.acm.org/10.1145/3003965.3003970*

Mazimpaka, J. D. and Timpf, S. (2017), 'How they move reveals what is happening: Understanding the dynamics of big events from human mobility pattern', *ISPRS International Journal of Geo-Information* **6**(1), 15.

Meijles, E., de Bakker, M., Groote, P. and Barske, R. (2014), 'Analysing hiker movement patterns using GPS data: Implications for park management', *Computers, Environment and Urban Systems* **47**, 44 – 57. Progress in Movement Analysis – Experiences with Real Data.
**URL:** *http://www.sciencedirect.com/science/article/pii/S0198971513000665*

Mennis, J. and Guo, D. (2009), 'Spatial data mining and geographic knowledge discovery—an introduction', *Computers, Environment and Urban Systems* **33**(6), 403 – 408.
**URL:** *http://www.sciencedirect.com/science/article/pii/S0198971509000817*

Miluzzo, E., Lane, N. D., Eisenman, S. B. and Campbell, A. T. (2007), CenceMe: Injecting sensing presence into social networking applications, *in* G. Kortuem, J. Finney, R. Lea and V. Sundramoorthy, eds, 'Proceedings of the 2nd European Conference on Smart Sensing and Context (EuroSSC 2007)', Springer-Verlag, Berlin, Heidelberg, pp. 1–28.
**URL:** *http://dl.acm.org/citation.cfm?id=1775377.1775379*

Mitas, L., Brown, W. M. and Mitasova, H. (1997), 'Role of dynamic cartography in simulations of landscape processes based on multivariate fields', *Computers & Geosciences* **23**(4), 437 – 446.
**URL:** *http://www.sciencedirect.com/science/article/pii/S0098300497000071*

Mokhtar, H. M. O. and Su, J. (2004), Universal trajectory queries for moving object databases, *in* 'Proceedings of the 5th IEEE International Conference on Mobile Data Management (Berkeley, CA, USA, 19-22 January 2004)', IEEE Computer Society, pp. 133–144.

Nanni, M. and Pedreschi, D. (2006), 'Time-focused clustering of trajectories of moving objects', *Journal of Intelligent Information Systems* **27**(3), 267–289.
**URL:** *http://dx.doi.org/10.1007/s10844-006-9953-7*

Orellana, D., Bregt, A. K., Ligtenberg, A. and Wachowicz, M. (2012), 'Exploring visitor movement patterns in natural recreational areas', *Tourism Management* **33**(3), 672–682.
**URL:** *http://dx.doi.org/10.1016/j.tourman.2011.07.010*

Orellana, D. and Renso, C. (2010), Developing an interactions ontology for characterizing pedestrian movement behaviour, *in* M. Wachowicz, ed., 'Movement-Aware Applications For Sustainable Mobility: Technologies And Approaches', IGI Global, Hershey, PA, USA, pp. 62–86.

Orellana, D. and Wachowicz, M. (2011), 'Exploring patterns of movement suspension in pedestrian mobility', *Geographical Analysis* **43**(3), 241–260.
**URL:** *http://dx.doi.org/10.1111/j.1538-4632.2011.00818.x*

Palma, A. T., Bogorny, V., Kuijpers, B. and Alvares, L. O. (2008), A clustering-based approach for discovering interesting places in trajectories, *in* 'SAC '08: Proceedings of the 2008 ACM symposium on Applied computing', ACM, Fortaleza, Brazil, pp. 863–868.
**URL:** *http://dx.doi.org/10.1145/1363686.1363886*

Pan, G., Qi, G., Zhang, W., Li, S., Wu, Z. and Yang, L. (2013), 'Trace analysis and mining for smart cities: issues, methods, and applications', *IEEE Communications Magazine* **51**(6), 120–126.
**URL:** *http://dx.doi.org/10.1109/mcom.2013.6525604*

Papadias, D. and Theodoridis, Y. (1997), 'Spatial relations, minimum bounding rectangles, and spatial data structures', *International Journal of Geographical Information Science* **11**(2), 111–138.
**URL:** *http://dx.doi.org/10.1080/136588197242428*

Pelekis, N. and Theodoridis, Y. (2014), *Mobility Data Management and Exploration*, Springer-Verlag New York.
**URL:** *http://dx.doi.org/10.1007/978-1-4939-0392-4*

Pettit, B., Perna, A., Biro, D. and Sumpter, D. J. T. (2013), 'Interaction rules underlying group decisions in homing pigeons', *Journal of The Royal Society Interface* **10**(89).
**URL:** *http://rsif.royalsocietypublishing.org/content/10/89/20130529*

Piorkowski, M., Sarafijanovic-Djukic, N. and Grossglauser, M. (2009), 'Crawdad data set epfl/mobility (v. 2009-02-24)', http://crawdad.org/epfl/mobility/20090224.
**URL:** *https://doi.org/10.15783/C7J010. Accessed on 5th October 2014*

Polous, K., Freitag, A., Krisp, J., Meng, L. and Singh, S. (2015), A general framework for event detection from social media, *in* F. Harvey and Y. Leung, eds, 'Advances in Spatial Data Handling and Analysis: Select Papers from the 16th IGU Spatial Data Handling Symposium', Springer International Publishing, Cham, pp. 85–105.

Praing, R. and Schneider, M. (2007), A universal abstract model for future movements of moving objects, *in* S. I. Fabrikant and M. Wachowicz, eds, 'The European Information Society: Leading the Way with Geo-information', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 111–120.

Randell, D. A., Cui, Z. and Cohn, A. G. (1992), A spatial logic based on regions and connection, *in* B. Nebel, C. Rich and W. Swartout, eds, 'Proceedings of the 3rd International Conference on Knowledge Representation and Reasoning', Morgan Kaufmann Publishers Inc., pp. 165–176.

Rattenbury, T., Good, N. and Naaman, M. (2007), Towards automatic extraction of event and place semantics from Flickr tags, *in* 'Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval', SIGIR '07, ACM, New York, NY, USA, pp. 103–110.
**URL:** *http://doi.acm.org/10.1145/1277741.1277762*

Reis, R. M., Egenhofer, M. J. and Matos, J. L. (2008), Conceptual neighborhoods of topological relations between lines, *in* A. Ruas and C. Gold, eds, 'Headway in Spatial Data Handling: 13th International Symposium on Spatial Data Handling', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 557–574.

Rigaux, P., Scholl, M. and Voisard, A. (2001), *Spatial databases with application to GIS*, Morgan Kaufmann.

Rokach, L. (2010), A survey of clustering algorithms, *in* O. Maimon and L. Rokach, eds, 'Data Mining and Knowledge Discovery Handbook', Springer US, Boston, MA, pp. 269–298.

Roquet, F., Wunsch, C., Forget, G., Heimbach, P., Guinet, C., Reverdin, G., Charrassin, J.-B., Bailleul, F., Costa, D. P., Huckstadt, L. A., Goetz, K. T., Kovacs, K. M., Lydersen, C., Biuw, M., Nøst, O. A., Bornemann, H., Ploetz, J., Bester, M. N., McIntyre, T., Muelbert, M. C., Hindell, M. A., McMahon, C. R., Williams, G., Harcourt, R., Field, I. C., Chafik, L., Nicholls, K. W., Boehme, L. and Fedak, M. A. (2013), 'Estimates of the southern ocean general circulation improved by animal-borne instruments', *Geophysical Research Letters* **40**(23), 6176–6180. 2013GL058304.
**URL:** *http://dx.doi.org/10.1002/2013GL058304*

Ross, S. M. (2014), *Introduction to probability and statistics for engineers and scientists*, 5th edn, Elsevier Academic Press, Burlington, MA, USA.

Safi, K., Kranstauber, B., Weinzierl, R., Griffin, L., Rees, E. C., Cabot, D., Cruz, S., Proaño, C., Takekawa, J. Y., Newman, S. H. et al. (2013), 'Flying with the wind: scale dependency of speed and direction measurements in modelling wind support in avian flight', *Movement Ecology* **1**(1), 4.

Salamat, N. and Zahzah, E.-h. (2012), 'Spatiotemporal relations and modeling motion classes by combined topological and directional relations method', *ISRN Machine Vision* **2012**(ID 872687).

Schlieder, C. (1995), Reasoning about ordering, *in* A. U. Frank and W. Kuhn, eds, 'Spatial Information Theory - A Theoretical Basis for GIS: Proceedings of COSIT '95 (Semmering, Austria, September 21–23, 1995)', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 341–349.

Sharker, M. H. and Karimi, H. A. (2014), 'Computing least air pollution exposure routes', *International Journal of Geographical Information Science* **28**(2), 343–362.
**URL:** *http://dx.doi.org/10.1080/13658816.2013.841317*

Siła-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U. and Fotheringham, A. S. (2016), 'Analysis of human mobility patterns from GPS trajectories and contextual information', *International Journal of Geographical Information Science* **30**(5), 881–906.
**URL:** *http://dx.doi.org/10.1080/13658816.2015.1100731*

Sklar, M., Shaw, B. and Hogue, A. (2012), Recommending interesting events in real-time with foursquare check-ins, *in* 'Proceedings of the Sixth ACM Conference on Recommender Systems', RecSys '12, ACM, New York, NY, USA, pp. 311–312.
**URL:** *http://doi.acm.org/10.1145/2365952.2366028*

Spaccapietra, S., Parent, C. and Spinsanti, L. (2013), Trajectories and their representations, *in* C. Renso, S. Spaccapietra and E. Zimanyi, eds, 'Mobility Data', Cambridge University Press, New York, pp. 3–22.

Spinsanti, L., Berlingerio, M. and Pappalardo, L. (2013), Mobility and geo-social networks, *in* C. Renso, S. Spaccapietra and E. Zimanyi, eds, 'Mobility Data', Cambridge University Press, New York, pp. 315–333.
**URL:** *http://dx.doi.org/10.1017/cbo9781139128926.017*

Tarroux, A., Weimerskirch, H., Wang, S.-H., Bromwich, D. H., Cherel, Y., Kato, A., Ropert-Coudert, Y., Øystein Varpe, Yoccoz, N. G. and Descamps, S. (2016), 'Flexible flight response to challenging wind conditions in a commuting antarctic seabird: do you catch the drift?', *Animal Behaviour* **113**, 99 – 112.
**URL:** *http://www.sciencedirect.com/science/article/pii/S0003347215004741*

Timpf, S. and Devogele, T. (1997), New tools for multiple representations, *in* L. Ottoson, ed., 'Proceedings of the 18th International Cartographic Conference (Stockholm, Sweden, 23-27 June 1997)', Swedish Cartographic Society, pp. 1381–1386.

Tominski, C., Schumann, H., Andrienko, G. and Andrienko, N. (2012), 'Stacking-based visualization of trajectory attribute data', *IEEE Transactions on Visualization and Computer Graphics* **18**(12), 2565–2574.

Vallet, D., Castells, P., Fernandez, M., Mylonas, P. and Avrithis, Y. (2007), 'Personalized content retrieval in context using ontological knowledge', *IEEE Transactions on Circuits and Systems for Video Technology* **17**(3), 336–346.

Van de Weghe, N. (2004), Representing and reasoning about moving objects: a qualitative approach, phdthesis, Ghent University.

Van de Weghe, N., Cohn, A. G., De Tre, G. and De Maeyer, P. (2006), 'A qualitative trajectory calculus as a basis for representing moving objects in geographical information systems', *Control and Cybernetics* **35**(1), 97–119.

Van de Weghe, N. and De Maeyer, P. (2005), Conceptual neighbourhood diagrams for representing moving objects, *in* J. Akoka, S. W. Liddle, I.-Y. Song, M. Bertolotto, I. Comyn-Wattiau, W.-J. van den Heuvel, M. Kolp, J. Trujillo, C. Kop and H. C. Mayr, eds, 'Perspectives in Conceptual Modeling: Proceedings of ER 2005 Workshops AOIS, BP-UML, CoMoGIS, eCOMO, and QoIS (Klagenfurt, Austria, October 24-28, 2005)', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 228–238.

Vandecasteele, A., Devillers, R. and Napoli, A. (2014), 'From movement data to objects behavior using semantic trajectory and semantic events', *Marine Geodesy* **37**(2), 126–144.
**URL:** *http://dx.doi.org/10.1080/01490419.2014.902885*

Von Groote-Bidlingmaier, C. and Timpf, S. (2012), Modeling and analysis of migration movement of pre-colonial cultures in the amazon basin, *in* H. Pundt and L. Bernard, eds, 'Proceedings of the 1st AGILE PhD School (Wernigerode, Germany, 2012)', Shaker Verlag, pp. 110–117.
**URL:** *http://openaccess.city.ac.uk/624/*

Wachowicz, M., Ong, R. and Renso, C. (2013), Tailoring trajectories and their moving patterns to contexts, *in* D. Vandenbroucke, B. Bucher and J. Crompvoets, eds, 'Geographic Information Science at the Heart of Europe', Springer International Publishing, Cham, pp. 285–303.

Wachowicz, M., Ong, R., Renso, C. and Nanni, M. (2011), 'Finding moving flock patterns among pedestrians through collective coherence', *International Journal of Geographical Information Science* **25**(11), 1849–1864.
**URL:** *http://dx.doi.org/10.1080/13658816.2011.561209*

Wang, F., Chen, W., Wu, F., Zhao, Y., Hong, H., Gu, T., Wang, L., Liang, R. and Bao, H. (2014), A visual reasoning approach for data-driven transport assessment on urban roads, *in* 'Proceedings of the IEEE Conference on Visual Analytics Science and Technology (Paris, France, 25 - 31 October 2014)', IEEE, pp. 103–112.

Yan, Z., Parent, C., Spaccapietra, S. and Chakraborty, D. (2010), A hybrid model and computing platform for spatio-semantic trajectories, *in* L. Aroyo, G. Antoniou, E. Hyvönen, A. ten Teije, H. Stuckenschmidt, L. Cabral and T. Tudorache, eds, 'The Semantic Web: Research and Applications; ESWC 2010', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 60–75.

Zhao, J., Forer, P. and Harvey, A. S. (2008), 'Activities, ringmaps and geovisualization of large human movement fields', *Information Visualization* **7**(3), 198–209.
**URL:** *http://dx.doi.org/10.1057/palgrave.ivs.9500184*

Zheng, Y. (2015), 'Trajectory data mining: An overview', *ACM Transactions on Intelligent Systems and Technology* **6**(3), 1–41.
**URL:** *http://dx.doi.org/10.1145/2743025*

Zheng, Y., Chen, Y., Li, Q., Xie, X. and Ma, W.-Y. (2010), 'Understanding transportation modes based on GPS data for web applications', *ACM Transactions on the Web* **4**(1), 1–36.
**URL:** *http://dx.doi.org/10.1145/1658373.1658374*

Zheng, Y. and Zhou, X., eds (2011), *Computing with spatial trajectories*, Springer, New York.