



HAL
open science

A conceptual and semantic modelling approach for the representation and exploration of human trajectories

Meihan Jin

► **To cite this version:**

Meihan Jin. A conceptual and semantic modelling approach for the representation and exploration of human trajectories. Geography. Université de Bretagne occidentale - Brest, 2017. English. NNT : 2017BRES0067 . tel-01591177v2

HAL Id: tel-01591177

<https://hal.science/tel-01591177v2>

Submitted on 31 Jan 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



université de bretagne
occidentale

UNIVERSITE
BRETAGNE
LOIRE

THÈSE / UNIVERSITÉ DE BRETAGNE OCCIDENTALE

sous le sceau de l'Université Bretagne Loire

pour obtenir le titre de

DOCTEUR DE L'UNIVERSITÉ DE BRETAGNE OCCIDENTALE

Mention : Géomatique

École Doctorale des Sciences de la Mer

présentée par

Meihan Jin

Préparée à L'Institut de Recherche de
l'École Navale(IReNav)

A Conceptual and Semantic Modelling Approach for the Representation and Exploration of Human Trajectories

Thèse soutenue le 18 septembre, 2017

devant le jury composé de :

Marie-Hélène de Sède

Professeur des Universités, Université de Franche-Comté / *Rapporteur*

Marius Thériault

Professeur des Universités, Université Laval / *Rapporteur*

Jean-Philippe Babau

Professeur des Universités, Université de Bretagne Occidentale / *Examineur*

Jianya Gong

Professeur des Universités, Université of Wuhan / *Examineur*

Danielle Ziebelin

Professeur des Universités, Université de Grenoble / *Examineur*

Christophe Claramunt

Professeur des Universités, IRENav / *Directeur de thèse*



Acknowledgments

This work has been carried out at the French Naval Academy Research Institute that has been very supportive the past few years. The main financial support of my thesis provided by the China Scholarship Council (CSC) is gratefully acknowledged. It is no easy journey pursuing a PhD. However, thanks to the people who offer aids and support, this journey is also filled with joy and pleasure. Whereas it would lead me too far to mention each name exhaustively, I would like to express my sincere gratitude to all those people who had helped and supported me in this three years' time. Please allow me to mention some of them representatively.

First of all, I wish to appreciate the admirable academic guidance and selfless support in life from my supervisor Professor Christophe Claramunt. Without him, I would not have the chance to come to France and get a real start of my thesis. He had offered the most he could to help me finish my PhD thesis. I feel extremely honored and lucky to have him as my supervisor, not only because of his great academic achievements but also his kindness and nobility.

I also have to thank my supervisors in China, Xingquan Liu and Jianya Gong, who support me and encourage me to pursue my research goal abroad. Especially, Professor Liu and his wife Mrs Liu have always treated me as their own child, since the start of my graduate study and became a student of Professor Liu. I deeply appreciate them for their wise guidance on my research and the kind care for me.

I wish to thank my dearest parents, who not only have offered me the opportunity to enjoy my lifetime, but also guide me to a direction of life journey with morality and righteousness. They have given me more than I can expect and always believe in me. No words are enough to express my gratitude to them.

Also I would like to thank Monsieur Hamdi for making my PhD journey a pleasant one. He has encouraged me every time when I confronted frustration, and shared ideas and suggestions on my research. I sincerely gratitude the inspiration, support, and joy he has brought to me.

Last but not least, I want to express my gratitude to my dear friends and colleagues, Wissame, Loïc, Ludovic, Mathieu, Marie and all the others, who have helped me and treated me with all their hearts. It is my great pleasure to meet and know them all.

Table of Contents

Acknowledgments.....	i
Table of Contents.....	ii
List of Figures.....	vi
List of Tables.....	x
CHAPTER I: Introduction.....	1
I.1 Problem Statement.....	2
I.2 Aim and Objectives.....	3
I.3 Approach Overview.....	3
I.4 Contributions.....	7
I.5 Thesis Structure.....	8
CHAPTER II: Related Work.....	10
II.1 Introduction.....	11
II.2 Human Travel Patterns.....	11
II.3 The Concept of Trajectory.....	14
II.4 Trajectory Data Model.....	17
II.4.1 Conceptual Models.....	18
II.4.2 Event-Based Models.....	22
II.4.3 Process-Based Models.....	24
II.4.4 Logical Models.....	26
II.4.4.1 Vector-Based Models.....	27
II.4.4.2 Raster-Based Models.....	28
II.4.5 Novel Approaches.....	30
II.5 Summary and Discussion.....	31
CHAPTER III: Travel Displacement Modelling.....	35
III.1 Modelling Research Objective.....	36

III.2 Modelling Principles.....	36
III.3 Travel Displacement Model.....	41
III.3.1 Trajectory Module.....	43
III.3.2 Transportation Network Module.....	46
III.3.3 Travel Pattern Module.....	48
III.3.4 Relationships between Modules.....	54
III.4 Discussion.....	56
CHAPTER IV: Travel Displacement Database and Data Analysis.....	58
IV.1 Database Implementation Objectives.....	59
IV.2 Travel Displacement Database Principles.....	59
IV.3 Travel Displacement Analysis.....	65
IV.3.1 Data Analysis Categories.....	65
IV.3.1.1 Semantic Level.....	65
IV.3.1.2 Temporal Level.....	66
IV.3.1.3 Spatial Level.....	66
IV.3.1.4 Spatio-Temporal Levels.....	67
IV.3.2 Data Analysis Functions and Queries.....	70
IV.4 Discussion.....	83
CHAPTER V: Implementation & Experimental Evaluation.....	85
V.1 Implementation Objective.....	86
V.2 Experimental Data Set.....	86
V.3 Computing and Software Environments.....	89
V.3.1 Development Environment.....	89
V.3.2 Database Environment.....	89
V.4 Framework Construction.....	90
V.4.1 Data Pre-Processing.....	91
V.4.1.1 Construction of the Transportation Network.....	91

V.4.1.2	Map Matching.....	99
V.4.2	Database Construction and Analysis.....	102
V.4.2.1	Data Structure Generation.....	102
V.4.2.2	Route Extraction Algorithm.....	105
V.5	Results and Discussion.....	107
CHAPTER VI:	Conclusion & Perspectives.....	110
VI.1	Outline.....	120
VI.2	Main Contribution.....	121
VI.2.1	Conceptual Modelling.....	121
VI.2.2	Logical Design.....	122
VI.2.3	Data Manipulations & Functions.....	123
VI.2.4	Implementation.....	123
VI.2.5	Travel Behaviour Exploration.....	124
VI.3	Perspectives.....	125
VI.3.1	Modelling Extensions.....	125
VI.3.2	Data Analysis Extensions.....	126
VI.3.3	Data Manipulation Extensions.....	126
VI.3.4	Decision Making & Travel Prediction.....	128
APPENDIX A:	Résumé étendu de la thèse.....	129
A.1	Introduction.....	130
A.2	Approche de Modélisation.....	133
A.2.1	Modèle de Mobilité.....	133
A.2.2	Module Trajectoire.....	134
A.2.3	Module Transportation Network.....	135
A.2.4	Module Travel Pattern.....	136
A.2.5	Relations entre Différents Modules.....	141
A.3	Analyse de Patrons de Déplacements.....	142

A.3.1 Interrogations de comportements de déplacements (Travel Displacement Queries).....	142
A.4 Implementation et Validation.....	144
A.4.1 Conception de Fonctions.....	145
A.4.2 Données Expérimentales de Trajectoires.....	147
A.4.3 Extraction des Trajectoires.....	148
A.4.4 Résultats.....	148
A.4.5 Evaluation des Performances.....	157
A.5 Conclusion.....	158
APPENDIX B: List of Publications.....	159
APPENDIX C: Short Curriculum Vitae.....	160
REFERENCES.....	162

List of Figures

Figure I. 1. Thesis' Scientific and Application Domains.....	6
Figure II. 1. Daily Prisms from Hägerstraand (1970).....	16
Figure II. 2. A simple space-time prism, from Wu & Miller 2011.....	16
Figure II. 3. A modelling approach for trajectories (Spaccapietra & Parent, 2008).....	19
Figure II. 4. An urban network database schema (Etches & Claramunt, 1998).....	21
Figure II. 5. Schematic example of an individual history.....	23
Figure II. 6. Relational schema for describing household and residential lifelines (Claramunt & Thérialt 2009).....	23
Figure II. 7. Typology of spatio-temporal processes (Claramunt and Thériault, 2015).....	25
Figure II. 8. Possible states for one object over a transition (Egenhofer 2000).....	26
Figure II. 9. Realm-based spatial data type as suggested by Güting and Schneider (1993).....	27
Figure II. 10. Relative view of a trajectory configuration (Noyon and Claramunt, 2007).....	28
Figure II. 11. One-dimensional object evolution in a 2D space (Hallot and Billen 2007).....	28
Figure II. 12. Spatial changes at time t_i displayed as a simplified map(a) and the corresponding event components(b). (Peuquet, 1995).....	29
Figure II. 13. Intersecting trip vectors with grids (Liu 2015).....	30
Figure III. 1. Trajectory example.....	42
Figure III. 2. General Schema of the Travel Displacement Model.....	43
Figure III. 3. Trajectory Module.....	44

Figure III. 4. Transportation Module.....	48
Figure III. 5. Travel Pattern Module.....	50
Figure III. 6. The relationship between Trajectory and Route.....	55
Figure III. 7. The relationship between Spatial Object Types.....	56
Figure IV. 1. Query Categories.....	69
Figure IV. 2. Function1 Ind_At_Instant.....	71
Figure IV. 3. Function 2 Ind_Traj_During_Period.....	72
Figure IV. 4 Function 3 Ind_Enter_At.....	73
Figure IV. 5. Function 4 Ind_Exit_At.....	73
Figure IV. 6. Function 5 Ind_Confine_In.....	74
Figure IV. 7. Function 6 Traj_MBR.....	75
Figure IV. 8. Function 7 Traj_Pass_District.....	76
Figure IV. 9. Function 8 Traj_Pass_Transport.....	76
Figure IV. 10. Function 9 Traj_Nearest_POIs.....	77
Figure IV. 11. Function 10 Traj_Direction.....	79
Figure IV. 12. Function 11 Traj_Co-direction.....	79
Figure IV. 13. Function 12 Ind_Find_Route.....	80
Figure IV. 14. Function 13 Traj_Fit_Route.....	81
Figure IV. 15: Function 14 Ind_Share_path.....	82
Figure IV. 16. Function 15 Trans_Avg_speed.....	83
Figure V. 1. Geolife Dataset.....	87
Figure V. 2 Processing and Construction Procedure.....	91
Figure V. 3. Blade touches Input.....	93
Figure V. 4. Blade Crosses Input.....	93
Figure V. 5. Multi-Intersection.....	95

Figure V. 6. Algorithm 2. Transportation Network Construction.....	98
Figure V. 7. Algorithm 3. Map Matching processing.....	101
Figure V. 8. Algorithm 4. Trajectory Construction Principles.....	103
Figure V. 9. Algorithm 5. Route Extraction Algorithm.....	105
Figure V. 10. Transportation segments regularly repeated by Li Lei in the morning and evening.....	108
Figure V. 11. Weekday versus weekend routes for a given individual.....	109
Figure V. 12. Geographical range of trajectories performed by individuals during a given time.....	110
Figure V. 13. Districts revealed by the start of the subway Line 10 trajectories....	111
Figure V. 14. Most popular individual routes.....	113
Figure V. 15. Busiest routes.....	113
Figure V. 16. POIs near the destination of some trajectories in the morning.....	114
Figure V. 17. The Travel Range Derived from Beihang University (Region marked as light blue, the points are the trajectories moving from the university).....	115
Figure V. 18. Li Ming's Trajectories passed by Road '三环'.....	116
Figure V. 19. Shared Paths of Li Ming and Wang Lei.....	117
Figure VI. 1. PhD Research approach.....	121
Figure A. 1 Modèle de représentation des mobilités	134
Figure A. 2 Segments de transportation régulièrement répétés par Li Lei le matin et le soir.....	149
Figure A. 3 Routes les week-ends et en semaine pour un individu donné.....	150
Figure A. 4 Couverture spatiale des trajectoires effectuées par des individus donnés pour un temps donné.....	151
Figure A. 5 Districts matérialisant des débuts de trajectoires de la ligne de métro 10.....	153

Figure A. 6 Routes les plus fréquentées.....155
Figure A. 7 Routes à plus forts trafic.....155
Figure A. 8 Points d'intérêts POIs proches de destinations en matinée.....156

List of Tables

Table II. 1. Overview of the spatio-temporal modelling approaches.....	33
Table III. 1. MADS spatial and temporal object types and icons.....	40
Table IV. 1. TransportationNode Table (Related to the object TransportationNode)	60
Table IV. 2. TransportationSegment Table.....	61
Table IV. 3. Individual Table.....	61
Table IV. 4. TrajectoryMain Table.....	62
Table IV. 5. TrajectoryPartTable.....	62
Table IV. 6. Route Table.....	63
Table IV. 7. POI Table.....	63
Table IV. 8. District Table.....	64
Table IV. 9. Structure of the Activity Table.....	64
Table IV. 10. Environment Constraint Table.....	64
Table IV. 11. Weather Constraint Table.....	65
Table IV. 12. Spatio-temporal Operators, adapted from Wu and Claramunt (2014)...	70
Table V. 1. Logical representation of the trajectory dataset.....	88
Table V. 2. Transportation mode logical representation.....	88
Table V. 3. Computer Environment.....	89
Table A. 1 Liste de Fonctions.....	146



Introduction

The research developed by this PhD research is oriented towards the semantic modelling of human travel displacements derived from a very large urban trajectory database. The motivation behind this research originates from the necessity to search for travel patterns that emerge from citizens acting in the city and the massive movement data generated. We develop a conceptual and modelling framework whose objective is to integrate, represent and analyse human trajectories and mobility choices. The final aim is to provide a modelling environment that will favour further analysis of travel displacements in the city. The framework includes a conceptual and logical travel displacement model that represents the travel habits of citizens moving in a given city. The whole approach has been implemented in an extensible spatial database system, experimented in the context of real transportation data and enriched by a series of query interface manipulations that illustrate the full potential of our framework.

PROBLEM STATEMENT

Cities can be considered as dynamic and complex environments whatever the scale of study. Cities encompass a lot of human activities and interactions whose exploration is far from being a straightforward task. In particular, modern large cities generate high displacements and traffic flows that still deserve the development of appropriate modelling frameworks in order to provide urban planners computing infrastructures that will favour further analysis.

Searching for and studying human travel behaviours have been a prevailing research issue since the 80s (Hanson & Huff, 1986; Jiang et al, 2012; Pentland & Liu, 1999). One of the main reasons behind this trend is that the prediction of individuals' travel behaviours is an essential component of transportation planning and policy analysis (Kitamura, 1988). Indeed, the emergence of modern computing and software environments for integrating and manipulating transportation and traffic data have been a major boost for the development of these research works. An important objective of the search for a better understanding of travel behaviour research is for instance to develop the capability to represent and predict how individuals respond to changes in their environments and how such responses are temporally and spatially correlated. However, the understanding of individuals' travel behaviours is still a non-straightforward task, as this not only involves decision processes, but also includes several semantics, spatial and temporal constraints that still require the development of appropriate database infrastructures. Therefore, the development of representation and computational frameworks that will favour the analysis and discovery of human travel behaviours are still considered as a research challenge.

In order to derive a better understanding at the local and aggregated levels of the way the city 'pulses' in space and time, the exploration should reveal how do people usually travel from origins to destinations; the specific routes followed in a specific day of the week and how these patterns aggregate at the spatial scale? These very few examples

of research questions should be answered when providing a system oriented to the representation and exploration of people's mobility patterns in an urban environment.

AIM AND OBJECTIVES

From the problem statement, this thesis is clearly oriented to the development of a modelling and computational framework that should help to categorize and characterize citizens' travel displacements, these providing novel opportunities for urban studies. In particular, the design of the modelling framework leads to the following objectives:

- to model the concept of human trajectories in an urban environment by considering the semantics, spatial and temporal dimensions;
- to design a trajectory database model, both at the conceptual and logical levels;
- to categorize the semantic, spatial and temporal data manipulation and queries that allow to explore the different patterns and properties that emerge from the modelling framework;
- to design and implement a data manipulation interface on top of a given Geographical Information System that facilitate the experimental validation of the whole modelling approach;
- to implement a series of representative queries aiming at discovering and analysing some travel displacement patterns derived from human trajectories in the city;
- to validate the whole approach in the context of real data derived from the city of Beijing in mainland China.

APPROACH OVERVIEW

The final aim of our research and approach is to provide a modelling framework whose objective is to represent and manipulate human trajectories, and search for some peculiar mobility choices as performed by citizens acting in a given city. The searching for and

analysis of humans' displacement deriving from a massive set of trajectories performed in an urban environment cover several scientific domains at different levels from Geographical Information Science (GIS) to spatio-temporal databases and a potential application domain oriented to urban studies (Figure I.1).

- **Geographical Information Science (GIS).** GIS is nowadays considered as a leading science for the representation and analysis of complex geographical and temporal phenomena in urban and natural environments. Geographical Information Science has been defined as the scientific discipline that studies data structures and computational techniques to capture, represent, process, and analyse geographic information (Goodchild 1992, 2004). One of the main GIS contributions when studying a given phenomenon is its ability to provide spatial, temporal and semantic abstractions that can then be manipulated by querying and analysis functions. In short, GIS provides several modelling and methodological resources to represent and analyse complex geographical and temporal phenomena, this fitting the context of the transportation and behavioural context of our application context
- **Spatio-temporal Databases.** Spatial database structures provide a privileged means for describing the metrical, topological, and attribute characteristics of objects (Armstrong, 1988). A spatio-temporal database embodies spatio-temporal concepts and simultaneously captures spatial and temporal aspects of data. It deals with geometries changing over time (Chen, 2008). According to Güting (1999) spatio-temporal databases have been widely applied to locations of objects moving over invariant geometry also known as moving object databases. The development of our modelling framework oriented to the representation of human displacements in the city is practically supported by trajectory data derived from GPS devices in the cities of Beijing and implemented for validation purposes on top of a spatio-temporal database. Such spatial database system is used to store and manage trajectory data, as well as semantic attributes associated to these trajectories. Several spatio-temporal database principles and functionalities have been explored and used by our work, as well as the development of specific data manipulation procedures, functions, index systems. Overall the spatio-temporal database environment provides a sort of

computerised manipulation and representation ‘engine’ that favours the implementation and validation of our framework.

- **Urban studies** Urban studies have been long oriented to the study of several patterns such as the exploration of human behaviours as referred to the array of every physical action and observable emotion associated with individuals, as well as the human race as a whole (Trivedi, 2017). Modelling human behaviour is to abstract the characteristic of human behaviour from real world to semantic expressions. Transportation patterns as they emerge from people’s actions in the city encompass several behavioural components whose study at the local and aggregated levels should provide complementary resources to study and understand the reasons behind such patterns. In particular, the different social roles of individuals should be considered when performing and generating some human activities in a given urban environment. In order to derive a better understanding of human displacements and mobility choices, there is still a need to design computational frameworks and database framework that will support the development of further data manipulation capabilities. Indeed, our objectives are not to develop a behavioural model that will be far beyond the scope of our research objective, but rather to design a semantic and spatio-temporal model and information infrastructure that will favour the exploration of mobility patterns in space and time.

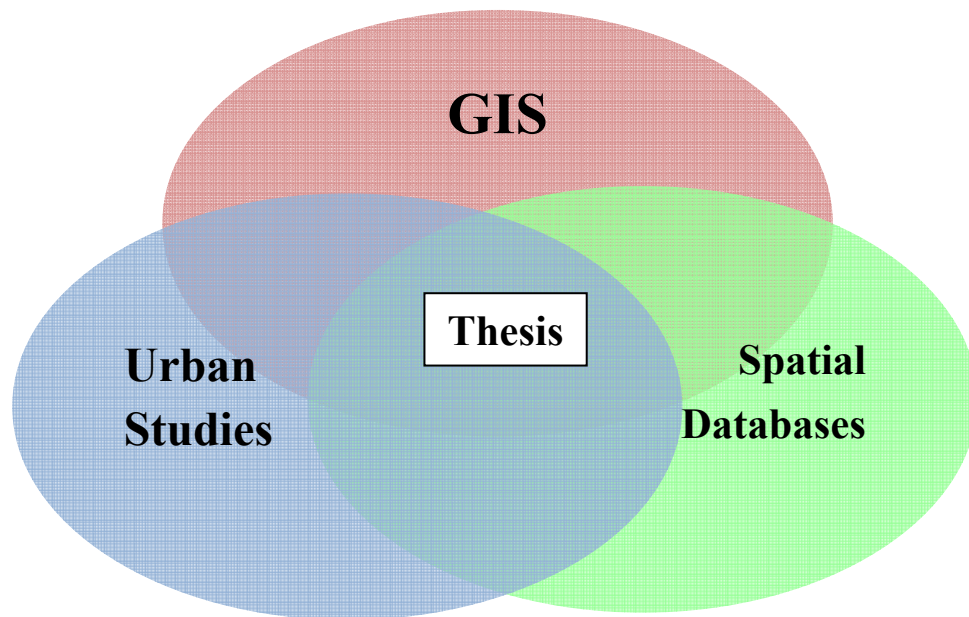


Figure I. 1. Thesis' Scientific and Application Domains

As previously mentioned, an important preliminary principle applied in this research is to develop our whole framework on top of a conceptual model that has proven to be appropriate for the representation of spatio-temporal data. This is the main motivation behind the choice of the MADS (for Modelling of Application Data with Spatio-temporal features) model developed by Spaccapietra & Parent (2006).

The conceptual model of travel displacements is developed as a foundation of trajectory data, trajectories being considered as spatio-temporal data produced from moving objects. An important starting modelling abstraction is the notion of *Individual* that performs some trajectories in the city. Several additional modelling abstractions are also considered by our modelling approach in order to take into account the different levels of semantics associated to trajectories. In particular, the underlying transportation network provides an important modelling component from which different entities are considered such as routes, transportation segments and transportation nodes. Additional entities are also considered such as individuals, points of interest, and districts.

A PostgreSQL/PostGIS travel displacement database is derived from a conceptual trajectory model. PostGIS is the spatial extension of PostgreSQL object-relational database, it provides spatial data types, spatial indexes and spatial data manipulation functions. These spatial data structures and functions provide a robust support to our implementation tasks. A series of interface development for the management and analysis of trajectory data has been designed, experimented and evaluated. Such interfaces are implemented as functions embedded in the PostgreSQL/PostGIS database. We developed a series of spatio-temporal and semantic queries on top of these interfaces, the objective being to analyse and discover citizens' travel displacements. Spatio-temporal and semantic queries are implemented and embedded within specific functions embedded in query language. This has the advantage of proposing specialized operations at the data manipulation level.

The experiment of our framework approach is performed on base of real movement data of people in the city. The dataset was collected by Geolife project (Microsoft Research Asia, 2008) with 178 users over a period of 4 years movement in Beijing City. It is recorded by GPS-equipped phones, and it contains 17,621 trajectories with a total distance of 1,251,654 kilometres and a total duration of 48,203 hours. The dataset is sufficient for testing the performance of our human travel trajectory framework, and it offers certain valuable information from people's displacements in the city.

CONTRIBUTIONS

The research developed in this thesis provides to our opinion several research contributions to the field of travel displacement modelling and geographical information science. The major findings of our work are as follows:

- A semantic and spatio-temporal model of human travel displacements based on urban trajectory data. The travel displacement model includes several modelling components encapsulated within specialized transportation network, trajectory and travel pattern data representation modules. The extent of the information represented in such modules is relatively large as spatial, temporal, semantics such as the potential factors that generate some travel decisions are

taken into account. The whole model has been implemented using the MADS modelling language.

- At the logical level, a complete travel displacement database has been derived from the conceptual level. Spatial object types, spatio-temporal object types and semantic object types identified by the conceptual model have been all created and generated in the PostgreSQL/PostGIS database. Specific data manipulation queries have been implemented as built-in functions at the interface and data query levels. This facilitates the expression of complex queries that combine spatial, semantics and temporal analysis.
- The experimental validation applied to real trajectory data shows evidence of the practicability of the travel displacement model and its database implementation. In particular, computing performances of complex queries have proven to be satisfactory as well as implementable. With such travel displacement approach, several human travel patterns can be explored, as well as a series of spatio-temporal relationships among multiple individuals have been extracted.

THESIS STRUCTURE

The remainder of this thesis is organized as follows:

Chapter II surveys relevant and related research works. A specific focus is made on current research oriented to the representation of human travel displacements, and how such displacements have been studied and analysed using preferable data modelling environments.

Chapter III introduces the main principles of our travel displacement model. The modelling approach is developed, as well as the reasons and motivation behind the choice of the MADS modelling framework further motivated. The main modelling modules of our framework are extensively presented, that is, the Trajectory Module, Transportation Module and Travel Pattern Modules.

Chapter IV develops the different components of the travel displacement database design. Spatial, temporal, semantic queries, as well as their combination, are extensively presented. We introduce the specialized functions whose objectives are to provide easy-to-apply typical queries aiming at the analysis of trajectory patterns. A series of representative examples are introduced in the context of real data.

Chapter V presents the implementation of the travel displacement model, including the dataset used for constructing the database and the implementation of the functions. As for complex functions, the algorithms implemented for their execution are detailed. And also in this chapter, the results of the experimental validation and queries applied to the cities of Beijing are illustrated. The computing execution performances of the queries are reported and the travel patterns and findings illustrated and discussed.

Chapter VI concludes the whole thesis and emphasizes the contribution of the research. Meanwhile, future research work and application perspectives are outlined.



Related Work

II.1	INTRODUCTION	11
II.2	HUMAN TRAVEL PATTERNS	11
II.3	THE CONCEPT OF TRAJECTORY	14
II.4	TRAJECTORY DATA MODEL.....	17
II.4.1	CONCEPTUAL MODELS	18
II.4.2	EVENT-BASED MODELS	21
II.4.3	PROCESS-BASED MODELS.....	24
II.4.4	LOGICAL MODELS	26
II.4.4.1	<i>Vector-based models</i>	27
II.4.4.2	<i>Raster-based models</i>	28
II.4.5	NOVEL APPROACHES.....	30
II.5	SUMMARY AND DISCUSSION	31

II.1 INTRODUCTION

This chapter presents a brief review of a series of research contributions oriented to the modelling of movement data and closely related to our research. This chapter is organized as follows. The next section presents related work on the development of conceptual models for the representation of human behaviour. Section 3 introduces the notion of trajectory and the motivation for designing a semantic and ontology-based model. Section 4 describes a series of modelling approaches for the representation of spatio-temporal and trajectory data. Finally, the last section categorizes and discusses the pro and con of previous research works, and motivates the preliminary principles of our modelling approach.

II.2 HUMAN TRAVEL PATTERNS

Researchers have been developing methodological and modelling frameworks for a long time to interpret and analyse human travel displacements (Alvares et al., 2007; Hanson & Huff, 1982; Hanson & Huff, 1985; Renso et al., 2013). According to Andrienko et al. (2008), a movement can be considered as a generalization of several geometrical and semantic primitives derived from individual trajectories, and that can be analysed at the global level in order to infer and generate additional knowledge. At the conceptual level, a trajectory can be represented as a user defined abstraction of the evolution of an object that is moving in space during a given time interval (Spaccapietra et al., 2008). At the application levels, comprehensive taxonomies have been introduced to characterize movement patterns for different domains such as the representation of animal trajectories and transportation patterns in the city or in the maritime area to mention a few examples (Dodge et al. 2008).

In order to discover the different properties that emerge from human trajectories, a diversity of algorithms have been proposed to characterize moving patterns. In the field of data mining, clustering is defined as the task of creating groups of objects that are similar to each other, while keeping separated those that are much different

(Nanni, 2013). Human activities can be aggregated into groups where every group encapsulates similar people's patterns. In a related work a spatio-temporal modelling approach has been developed for the analysis of mobility patterns at the city level (Thériault et al., 1999). Several modelling and group behaviour explorations have been introduced to characterize the behaviour of entity sets at different levels of abstraction. Modelling patterns and trends when observing human trajectories can be also observed using the notion of similarity that can be commonly evaluated using a measure of semantic distance (Parent et al., 2013). For example, Laube introduced several algorithms to detect movement patterns of dynamic objects, as derived and reflected by a matrix of motion attributes (Laube, 2005). The maximal length of the cumulated distances to the center is the primary constraint used. Another clustering algorithm oriented to the representation of human trajectories in an urban network has been proposed by Buchin et al. (2011). This approach is based on the Fréchet distance derived from two sub-trajectories computed under a constant time, so as to discover a common movement pattern of a group of entities. In the work developed by Hung & Peng (2009), trajectories are modelled as Probabilistic Suffix Trees (PST), and a clustering algorithm is based on a distance function applied to the PST. The K^{th} distance derived from the K-Nearest Neighbor (KNN) (Cover & Hart, 1967) theory can also act as a foundation of a clustering algorithm (Ong et al., 2010). Aside from the identification of clusters when analysing humans' movements, specific patterns across multiple moving entities can be also detected.

Generally, movement patterns include any recognizable spatial and temporal regularity or any interesting relationship in a set of movement data, whereas the proper definition of 'pattern interestingness' depends on the application domain (Dodge, 2008). Both the Lagrangian and Eulerian methods can be used to model movement (Goodwin et al., 2006). The Lagrangian method is usually applied to a moving object as it moves along a trajectory through space and time (Merki & Laube, 2012). The Eulerian method focuses on specific locations fixed in space and observes objects passing by. For instance, discovering random human walks in large movement patterns is another direction that has been explored by Gonzalez et al.

(2008). In their work, random walk methods such as the Lévy flight are used for unveiling the random movement of human beings.

In many applications, it appears that human beings tend to perform routine movements, although in many cases the detection of uncommon patterns is surely another important direction to consider. It is reasonable to assume that humans perform a high proportion of actions regularly as many objectives, obligations and constraints of humans' abilities can be considered as relatively stable (Schlich & Axhausen, 2003). In particular, Schlich showed that on average similarity patterns of human movements over a relatively long period (e.g., 6 weeks) reflected some significant trends. This is particularly reflected over day-to-day analysis especially on working days. Similar conclusions are also drawn by other researchers (Wakamiya et al., 2015). Moving patterns over weeks, as well as in space, has been highlighted in several Canadian and European cities (Axhausen et al., 2002; Buliung et al., 2008; Susilo & Axhausen, 2007). Noteworthy, it also appears that patterns over weekends are rather unstable in space (Srivastava & Schönfelder, 2003; Susilo & Kitamura, 2005). It is also believed that the diversity and repetition of the patterns of humans' movement in the city is plausibly due to the individual and household behavioural responses to divergent institutional constraints and potential differences in the spatial organization of cities and regions (Buliung et al., 2008).

While previous works have proven the existence of routine behaviours in human's movement in urban areas, it is still necessary to quantify such patterns as well as outlier behaviours, in order to derive a better understanding at the local and aggregated levels of the way the city 'pulses' in space and time. Such exploration should reveal how people usually travel from origins to destinations. The main objective of our research is to address this issue at the conceptual and implementation levels, and where the objective is to introduce an appropriate trajectory data model that favours the integration of real trajectory data sets and further analysis. The rest of this chapter focuses on the concept of trajectory and existing trajectory data models as introduced by current related work.

II.3 THE CONCEPT OF TRAJECTORY

The phenomenon of movement occurs whenever an object possesses different positions in space at different times (Galton, 1995). We live in a dynamic world, so movements happen everywhere, whatever the level of abstraction considered. It is an indivisible property in human life. People are in movement all time: they go to the kitchen for breakfast in the morning, they go to work, moving from one place to another etc. Whenever a change of position happens, there is a movement.

Thanks to many emerging and booming location-based and telecommunication technologies, a huge range of positioning devices are nowadays available, and that capture real-time position data of moving objects with relative high precision. For example, there are about 2 million cars on the road network of Beijing every day. If every 15 seconds the position of each car with equipped with a GPS device is recorded, this generates about 480 million records every hour! These huge amount of moving data gained by these location-based devices can offer valuable data and even valuable information if appropriate data structures and spatial data mining approaches can be successfully developed and applied. In fact, the large amount of data sources available increases the range of opportunities for geographical data analysis and mining, and this for many application domains. However, there is still a need to design and develop appropriate spatial data models as well as manipulation capabilities to take a full advantage of these new emerging big geographical data. Over the past few years, many researchers from the fields of GIS, database, data mining etc. have made great efforts in moving data management and analysis (Erwig, 1999; Guting., 2000; O’Sullivan, 2005; Spaccapietra et al., 2007; Zhene, 2009, Claramunt & Ray 2017).

A more precise definition of movement relates to the change of a physical position of a given entity with respect to some reference system (Andrienko, 2008). To complement the primitive spatial attribute of a given position, a movement abstraction should also take into account the time when a moving object leaves and reaches some locations. Moreover, a sequence of movement can be considered as a movement track, that is, a temporal sequence of spatio-temporal positions

(Andrienko, 2013). Furthermore, the successive segments of the movement track of a moving object generate the concept of trajectory. In a more mathematical way, a trajectory can be defined as the path that a moving object follows through space as a function of time (Lee & Krumm, 2011). For example, a taxi on the road can be regarded as a moving object. The action of the taxi moving from a given location, and picking up customers to their destination represents the movement of that taxi. The route taken from a given place modelled as a punctual primitive in a two-dimensional and/or network space, and from an origin at a given time, and to another place similarly modelled can be seen as a movement track,. Moreover, the different stops performed for picking up customers can be also considered as part of this movement track then considered as a trajectory. In order to understand the driving patterns of a given taxi driver for different periods, the segments of the movement tracks of that taxi during can be seen as a trajectory. Overall, both time and space play significant roles when addressing and formalizing the concepts of movement and trajectory. Mathematically, time can be considered as a continuous set with a linear ordering and a metric between instants that reflect instantaneous positions, or intervals that reflect movement durations. Physical time is not only a linear sequence of moments, but it includes inherent cycles resulting from the Earth's daily rotation and annual revolution (Andrienko, 2008).

Time Geography introduced by Hägerstrand (1970) combines time and space as a whole in an integrated modelling approach. A thorough concept of time-space is first introduced, that is generally considered as one of the early attempts oriented to the representation of movement and activities in geographical spaces. Overall, the modelling framework of Time Geography made a significant contribution to the modelling of spatio-temporal activities by also taking into account several spatial and temporal constraints related to movement capabilities. As a sort of precursor of time space research in GIS, Time Geography also provides some fundamental concepts for the integration of time within GIS, then providing some basic but important abstractions for a successful modelling of movement data. More recently, Miller (2009) applied Time Geography principles to design a model in which trajectories are modelled continuously as space-time paths, and where the concept of

prism is apply to identify possible movements as characterized by minimum cost curves through an inverse velocity field rather than straight line segments through a uniform velocity plane.

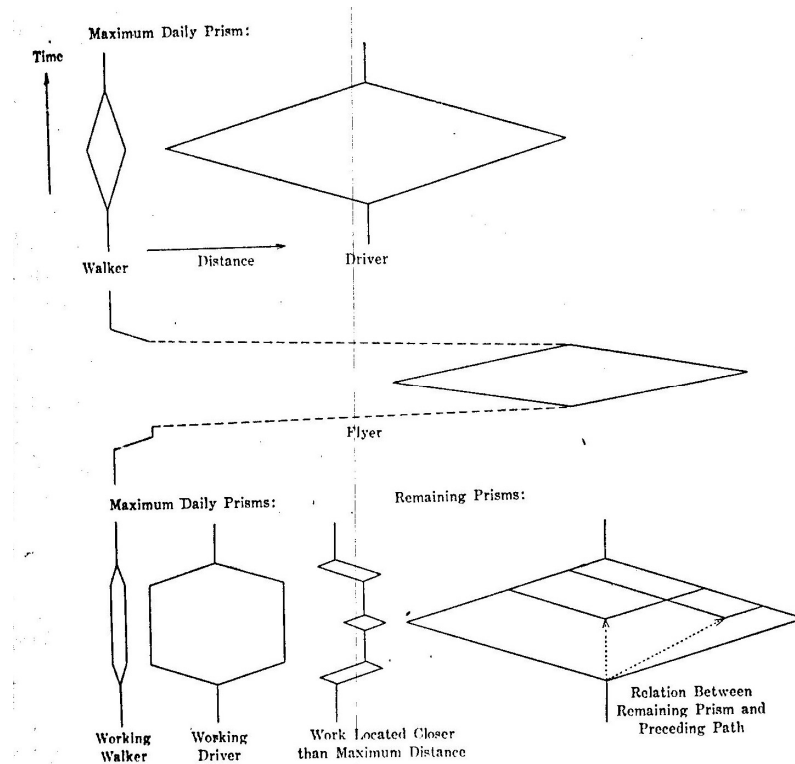


Figure II. 1. Daily Prisms from Hägerstrand (1970)

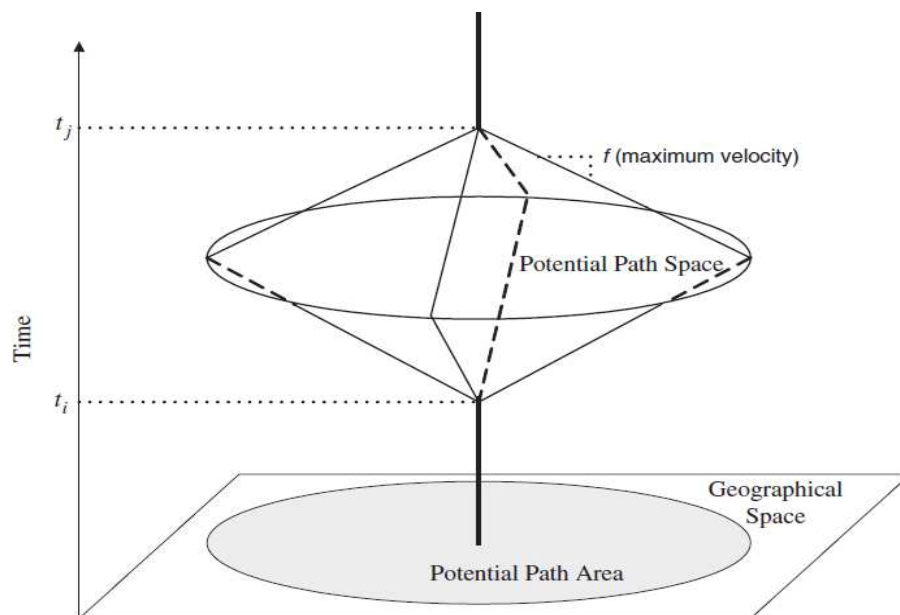


Figure II. 2. A simple space-time prism, from Wu & Miller 2011

II.4 TRAJECTORY DATA MODEL

Raw data generated by movement tracks can be considered as an implicit representation of several trajectories (Parent et al., 2013). Raw trajectories are particularly suited for tracking the successive locations of a moving object, computing statistics and deriving patterns and outliers. But as O’Sullivan noted (2005), any formal representation of geographical phenomena should take into account the different properties associated to such phenomena. In order to dig and extract valuable information from trajectories, raw trajectories only are not enough. In order to provide a more comprehensive representation of the semantics of a given trajectory, there is a need for a sort of ontology of the represented trajectory close to the domain and application of interest. This is particularly appropriate when applying GIS to environmental, social and human sciences that often associate the concept of trajectory to many additional semantics properties that help to characterize the many specific properties and characteristics of a primitive representation.

In fact, time when associated to space provides some additional crucial dimensions when it comes to fit people and things together, and when analysing the patterns derived from some socio-economic systems, so undergo long-term changes and events can be then potentially analysed if modelled appropriately. In order to mitigate the gap between human geography and information science, a recent proposal has introduced a semantic and ontological model of a trajectory (Yan & Spaccapietra, 2008). These authors developed a self-contained top-down semantic trajectory model, which includes both spatio-temporal features and semantic trajectory units.

Another significant aspect for the development of trajectory data management arises from the prominent role and prevalence of the database level. As databases, as well as geographical databases, are not only relatively mature technologies, but also well-developed frameworks at the conceptual and logical levels, a successful development of a modelling approach for movement and trajectories should indeed consider the database dimension as an indispensable one. For instance, many spatial

database systems such as Oracle spatial, PostgreSQL/PostGIS, SQLServer Spatial etc. provide several implementation capabilities worth being explored when considering possible implementations.

II.4.1 Conceptual Models

Nowadays, several researchers have made great progress in the development of conceptual spatio-temporal representations including ontological and semantic trajectory data models (Allen 1995, Brakatsoulas 2004, Li & Claramunt, 2007; Octavio, 2007, Noyon & Claramunt 2007, Wu & Claramunt 2014, Zheni & Claramunt 2015). A conceptual data model provides a high-level data structure that makes the connection between users' requirements and the systems design level (Gemino 2005). Kung & Solvberg (1986) stressed that conceptual models help to build the bridge between analysts and users, analysts and designers. A conceptual model plays an important role regarding trajectory modelling. For example, Hallot & Billen (2007) proposed a conceptual model of topological configurations in which in a one-dimensional space (i.e., typically a road considered as a spatial frame of reference) and where topological configurations between moving objects are categorized. Spaccapietra & Parent (2008) designed a spatial data model that has proven to be appropriate for several application domains such as for the study of bird migrations or transportation patterns in the city. This semantic model facilitates for example the search for directions and destinations of the movement and migrations of a series of birds, then showing how a well-structured semantic model of trajectories can provide a valuable framework to study a given phenomenon of interest.

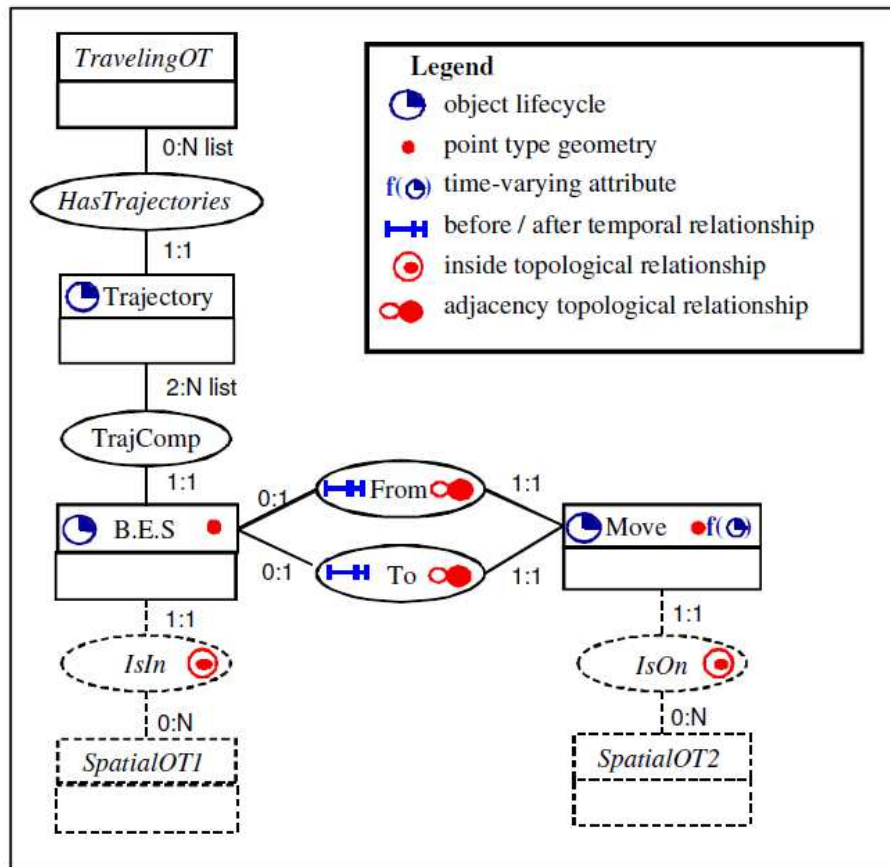


Figure II. 3. A modelling approach for trajectories (Spaccapietra & Parent, 2008)

One of the main advantages of such conceptual modelling approaches is that it facilitates the description of the semantic of a given application according to some modelling principles previously defined. For example, the database schema proposed by Etches & Claramunt (1998) provides a graphical view of a traffic database and facilitates the development of traffic management systems (Valsecchi & Claramunt, 1999).

Overall, the necessary elements to construct a trajectory database should include a conceptual data model, a logical data model and a physical data model. An example of semantic data model for trajectories is given by the work of Zheni & Claramunt (2009) that includes trajectory data types, spatial and temporal operations, and a sort of foundation for the database integration and manipulation of trajectory data.

In order to further implement a given conceptual model at the database level, several constructs can be applied. In particular, the notion of Abstract Data Type (ADT) has been initially suggested to model complex data types (Zheni & Claramunt, 2009). Su et al. (2001) represented a trajectory as an ADT this favouring its manipulation at the primitive query language level. Zheni et al. (2009) introduced an ADT to model a temporal trajectory data type, and additional semantic data operations. This semantic model is relatively generic and can be applied to different application domains. Erwig & Güting (1999) also proposed a logical model for moving points that can be directly mapped to a spatial database. The model is based on spatio-temporal data types of moving points and moving regions are identified, and both generic operations (e.g., “start”, “stop”) and derived spatio-temporal operations (e.g., “traverse”, “visits”) are defined as logic expressions. Moreover, such data model is developed in to a relatively comprehensive model which can efficiently manage moving object in multiple application environments including public transportation networks, indoor and outdoor spaces, etc. (Xu & Güting, 2013). Basic data types for trajectories are identified within a hierarchy and can act as primitives for application-based abstractions. Furthermore, basic data manipulation operations can be extended, towards for instance specific operations for temporal, spatial and spatio-temporal management. Such generic trajectory representation can be applied to many different real world environments.

Several additional research works have also made a contribution to the database modelling of trajectories. Vazirgiannis & Wolfson (2001) developed a trajectory data model oriented to the urban transportation domain. Brakatsoulas’s (2004) moving data model is also based on a road network, and is designed on top of a two-dimensional space and considers time as the third dimension. The movement of a given vehicle is registered and its semantics and properties integrated in a conceptual model. The model can support feature extraction for pattern discovery. A few other moving data models are designed on top of a road network (Yan & Parent, 2013). Indeed, a road network plays an important role in many trajectory models and offers several application perspectives.

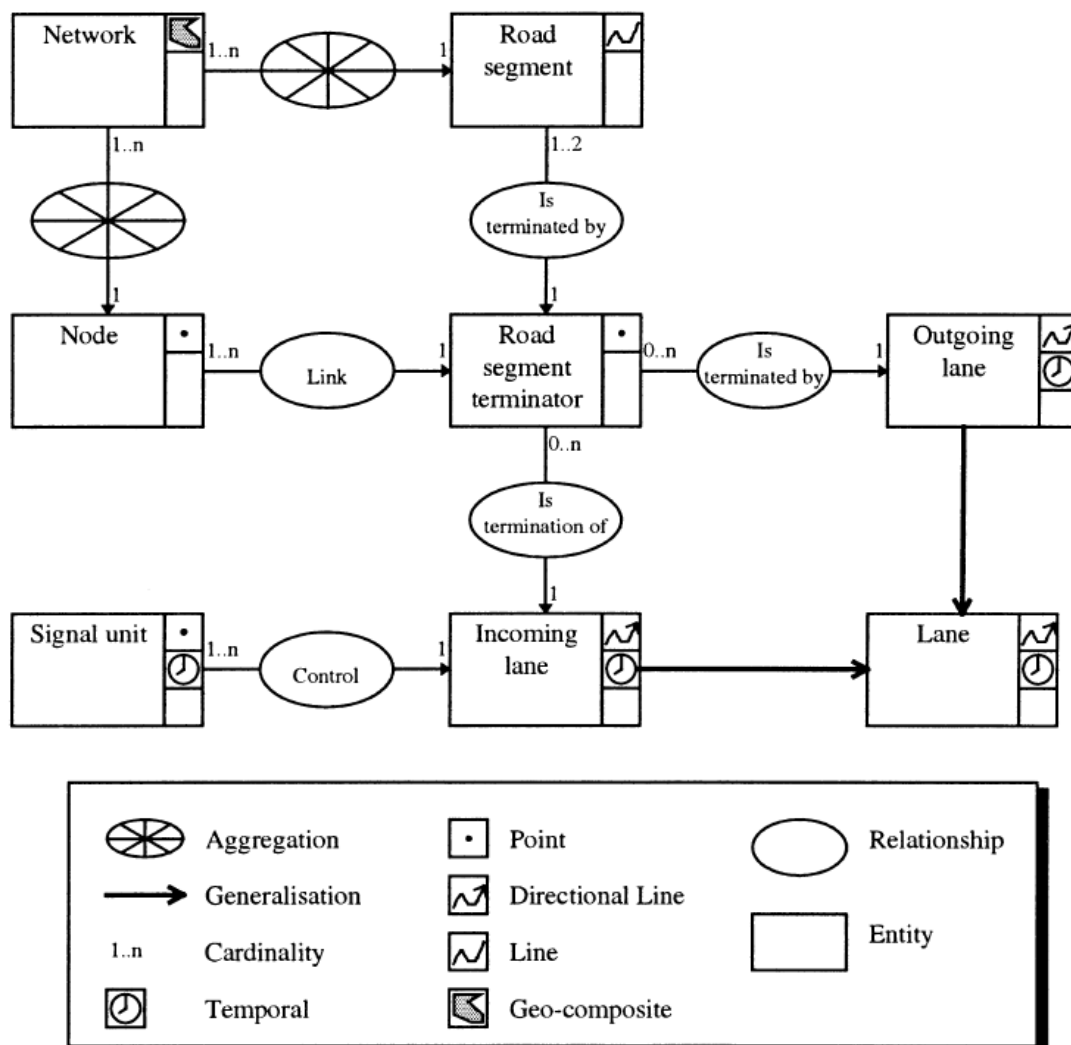


Figure II. 4. An urban network database schema (Etches & Claramunt, 1998)

Considering a spatio-temporal database as a valuable mediator to represent, manage and manipulate trajectories is nowadays accepted by both scientific and users' communities. One of the advantages of the ADT approach is that the semantics of any trajectory can be embedded in a given database system this offering multiple manipulation capabilities.

II.4.2 Event-based models

A significant amount of spatio-temporal modelling research has been conducted over the past few years towards event-based models (Allen, 1995; Worboys, 1995; Claramunt & Thériault, 2002). A survey published by Allen (1995) shows that

events are useful abstractions, and should be modelled as entities in order to further explore the notion of causality. According to Galton (1995), a theoretical basis is necessary to describe events regarding particularly the representation of the conditions for their occurrence. It introduces a framework for formalizing our “common-sense” knowledge of the real world. Allen et al. (1995) also introduced the concept of causality as a modelling abstraction that can be integrated into a temporal GIS, and developed a conceptual data model using an entity-relation formalism. This modelling approach identifies the specific elements that can contribute to the search of a causal theory, and the relationships between them.

Story & Worboys (1995) introduced a spatio-temporal model that considers events as a main primitive object associated to spatial, temporal and thematic attributes defined as basic components of that object abstraction. Claramunt & Thériault (1995) proposed an event-based conceptual model in which time is regarded as a complementary dimension that favours the representation and qualification of processes, thus favouring extensibility and manipulation at the query level. Such event-based modelling approach has been in particular applied to the representation of geo-lifelines (Claramunt & Thériault, 2002). Another example is the spatiotemporal scheduling framework introduced by Stewart et al. (2013). This framework is based on an ontological approach that identifies key process-based and event-based classes as well as the relations that connect them. These classes are considered as either continuants modelling real-world entities or occurrents capturing dynamics and events. The framework links a domain ontology (e.g., campus events) with an application ontology (e.g., scheduling), and a task ontology which allows individuals to plan and manage their schedule on a daily basis.

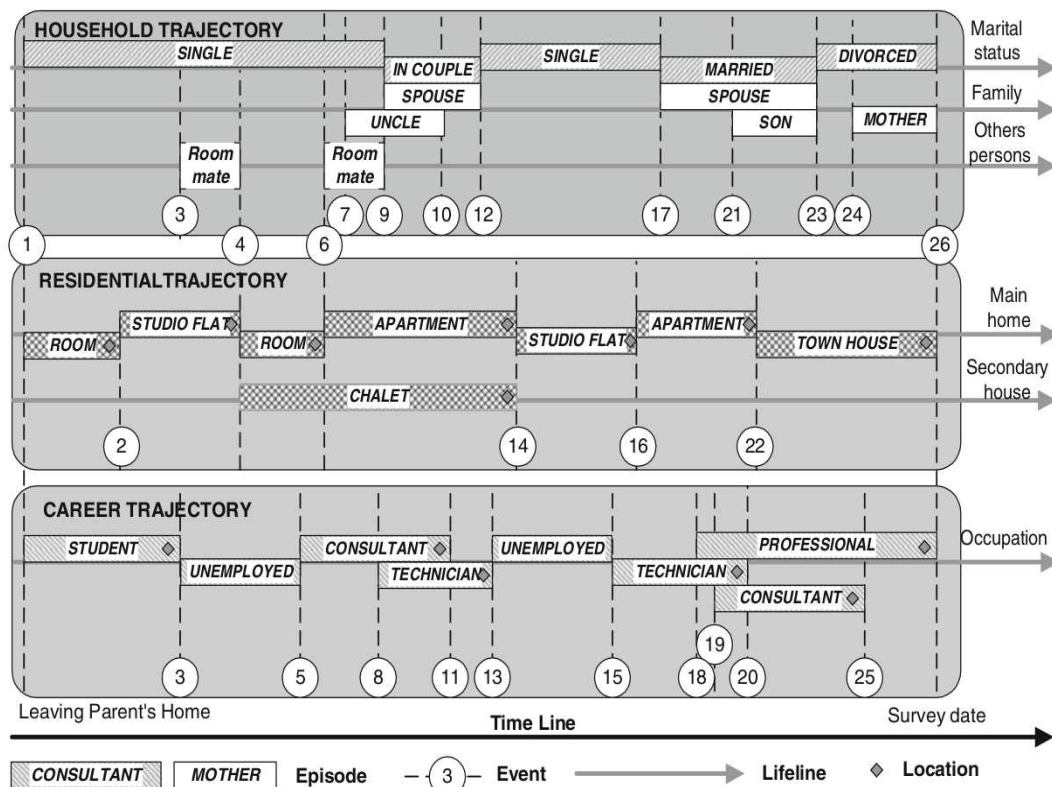


Figure II. 5. Schematic example of an individual history

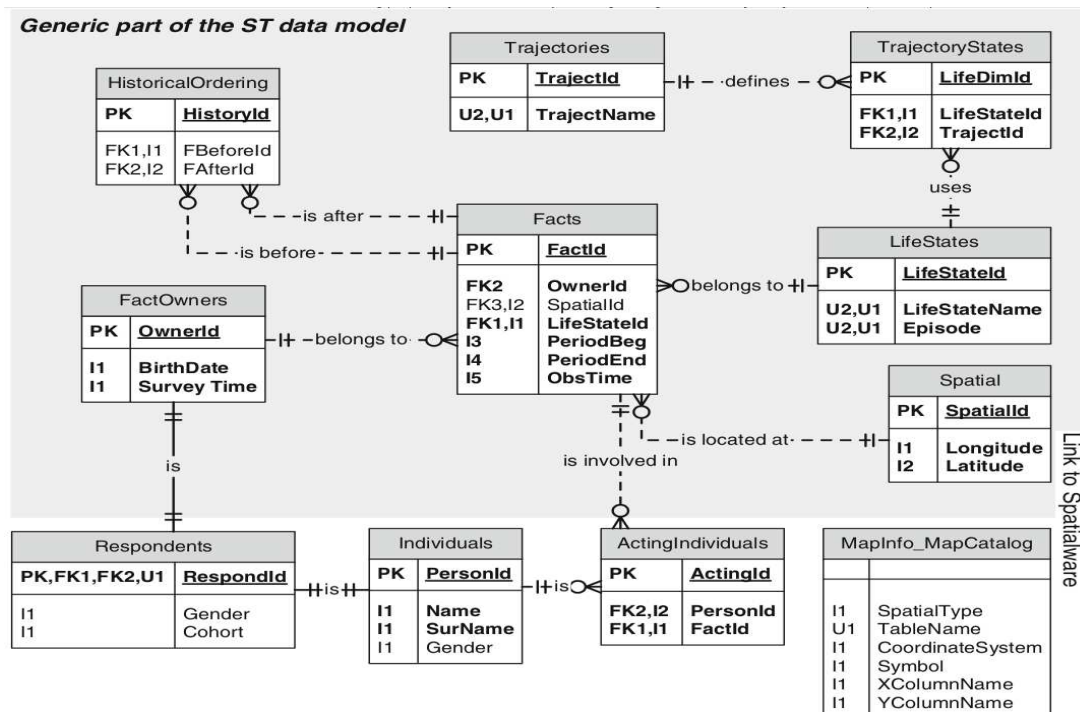


Figure II. 6. Relational schema for describing household and residential lifelines (Claramunt & Thériault 2009)

Other researcher works introduced a series of sematic models closer to some of the concepts suggested by Time Geography, where trajectories are represented as a series of movements and stops. For instance Alvares et al. (2007) developed a conceptual trajectory model where a series of algorithms are developed to extract stops and moves and moving patterns. Under similar principles several researches have designed trajectories models based on stops and moves (Spaccapietra 2007).

For instance, in the work of Yan and Parent (2008), a trajectory ontology is proposed and constructed on top of a road network. The trajectory ontology includes geometric, geography, and application-domain modules. The geometric module integrates the spatial, temporal, and spatio-temporal properties including trajectories. The geography and the application domain modules are rather application dependents, and describe the geographic and the thematic objects of the application. Fileto (2014) introduced a collection of modelling constructs to semantically describe movement data, and to support further analyses. These constructs include semantic trajectories, stops, and movements within a hierarchy.

Overall, event-based data models can be applied to many spatio-temporal phenomena, and where movements are associated to notions of starts, turnings, stops, as well as integrated within application-based ontologies.

II.4.3 Process-based models

Another modelling approach oriented to the representation of spatio-temporal data is based on the concept of process. The purpose is to categorize processes as they happen in geographical spaces, and to semantically qualify changes occurring in urban and natural environments (Claramunt et al., 1997). Such process-based models fill the conceptual gap between events that happened and the entities involved in such phenomena. The process categorization developed makes a difference between life and basic motion processes, and can be further enriched by additional spatial and temporal constraints to derive more complex and aggregated processes (Claramunt & Thériault, 2015).

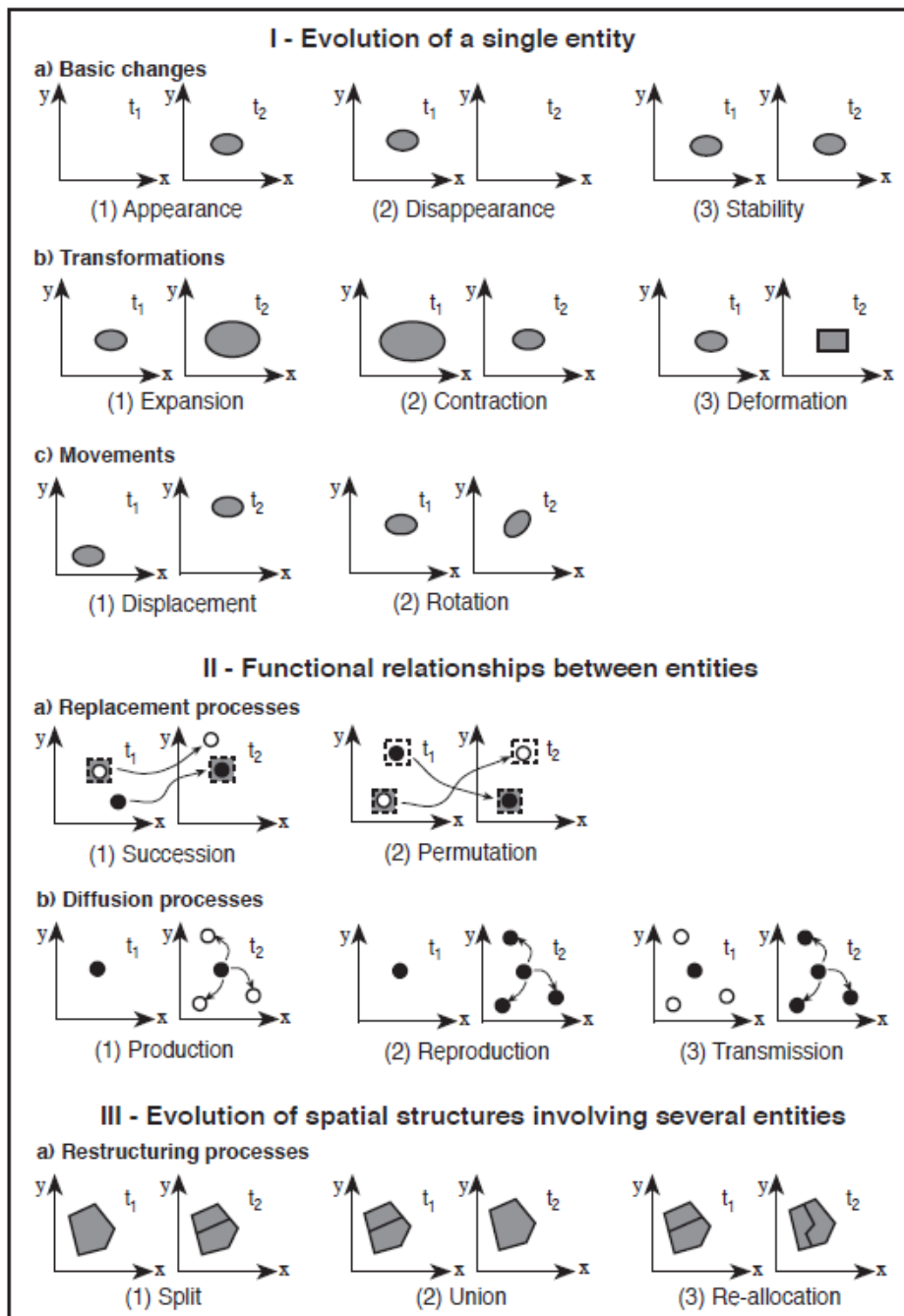


Figure II. 7. Typology of spatio-temporal processes (Claramunt & Thériault, 2015)

Claramunt & Thériault (1998) further introduced a set of design patterns for spatio-temporal processes. An integrated taxonomy of spatio-temporal processes is illustrated in Figure II.8. Hornsby & Egenhofer (2000) developed a spatio-temporal

knowledge representation based on the explicit expression of processes and changes oriented to geographic phenomena.

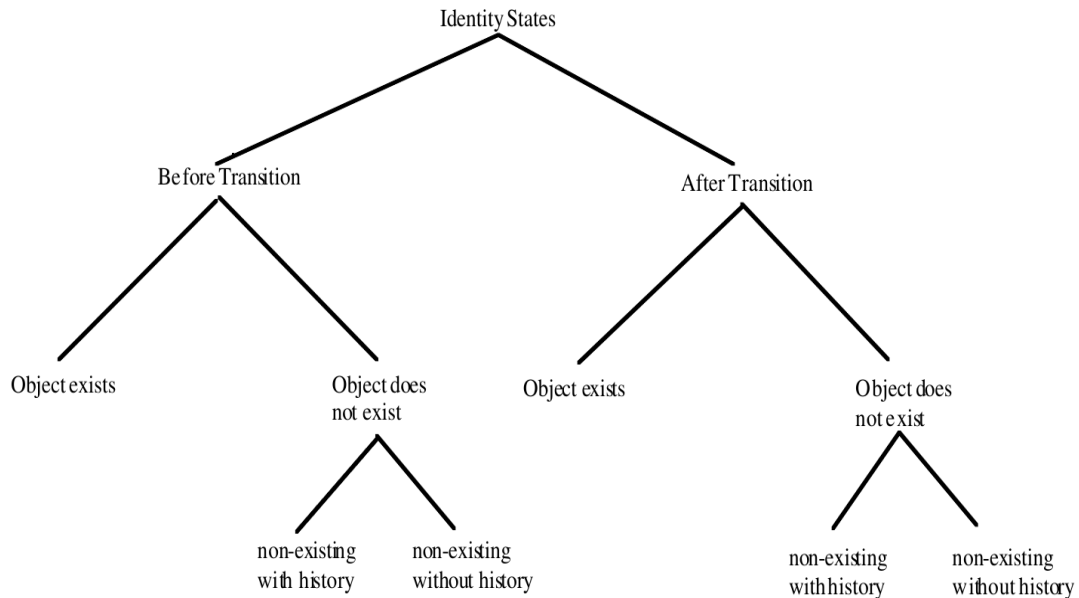


Figure II. 8. Possible states for one object over a transition (Egenhofer 2000)

II.4.4 Logical models

Logical data models are essential in information systems as they bridge the gap between the conceptual and implementation levels. Logical data models are generally materialized by either diagrammatic, relational or object-oriented languages and can even use graph representations to model evolving entities. One example of such logical models is the Realm model proposed by Güting & Schneider(1993). Spatial data is materialized in a spatial realm made up of a set of points and non-intersecting line segments as a sort of discrete-oriented representation of space as illustrated in Figure II.9.

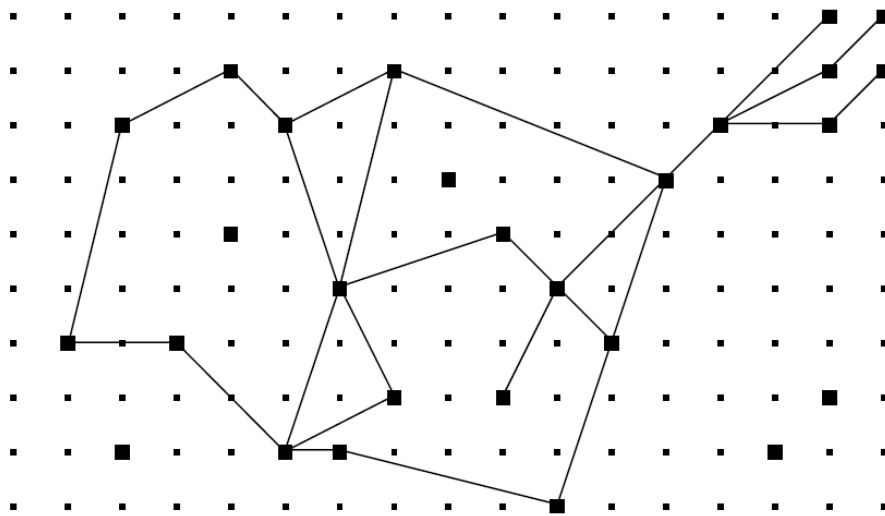


Figure II. 9. Realm-based spatial data type as suggested by Güting & Schneider (1993)

II.4.4.1 Vector-based models

One of the logical data models most often applied in geographical space is based on the so-called vector model (Hallot, 2007; Noyon, 2007; Miller, 2009; Liu, 2014). The vector-based data model is particularly adapted to represent spatio-temporal data such as trajectories. A logical data model designed as a relative representation for trajectories has been introduced by Noyon et al. (2007), in which the trajectory of a given mobile entity is defined by its relative speed distance according to a reference entity. This method is well adapted to capture the semantic of the movement of an entity with respect to a second one considered as an entity of interest. Liu (2014) suggested a vector-based model based on Time Geography concepts and a preliminary contribution of Miller & Bridwell (2009). This model can be applied to estimate the density of similar movements, and in order to extract moving patterns. Overall, vector-based models provide efficient frameworks to capture the spatial semantics of geographical phenomena, but they should be closely associated to the conceptual level and additional descriptive data in order to provide a more complete representation framework.

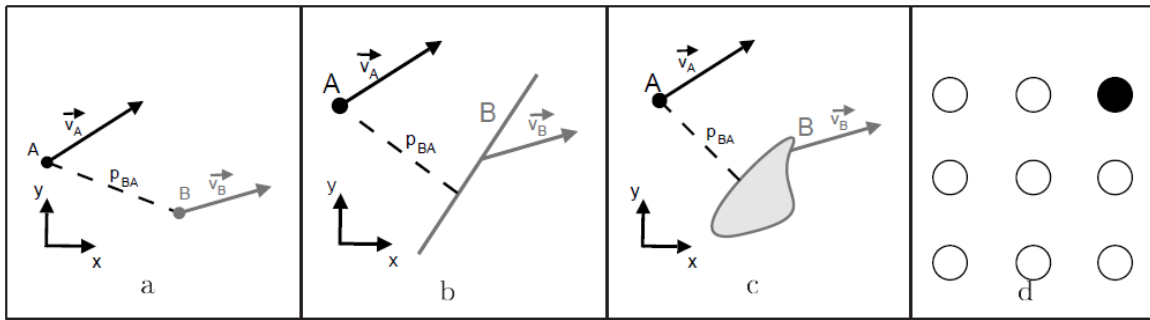


Figure II. 10. Relative view of a trajectory configuration (Noyon & Claramunt, 2007)

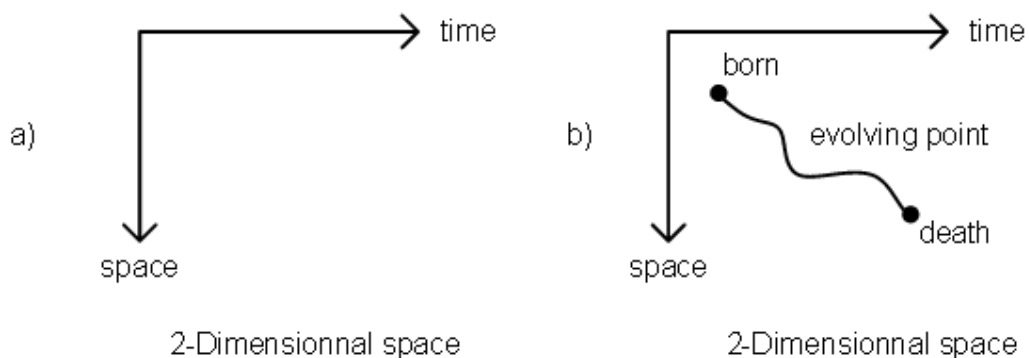


Figure II. 11. One-dimensional object evolution in a 2D space (Hallot & Billen 2007)

II.4.4.2 Raster-based models

Another approach that can be applied to represent trajectories, although less intuitive, is the raster model roughly defined as a representation, by most often a regular grid, of a continuous phenomenon. For instance, Peuquet (1995) suggested an Event-based Spatio-Temporal Data Model whose objective is to represent change patterns over time. Changes identified at a given time are represented at the primitive levels and related to cell locations. In other words, locations associated to some common semantic values are stored together as members of a same component. Another example is the one of Cellular-Automata (CA) models that have been widely used to model and replicate land-use evolution using combination of stochastic evolution rules, probabilities and comparison to observed data (Wu, 2002; Menard, 2005). These methods when applied to possible evolution scenarios can be cross-compared with projected land demand. Raster-based approaches can be further extended towards three-dimensional representation frame where the third dimension

represents the temporal dimension, and also combined with the vector-based model (Liu, 2015). The model is used to estimate a population demand vector field using a vector kernel density estimated from observed trajectories of a sample population (Figure II.13).

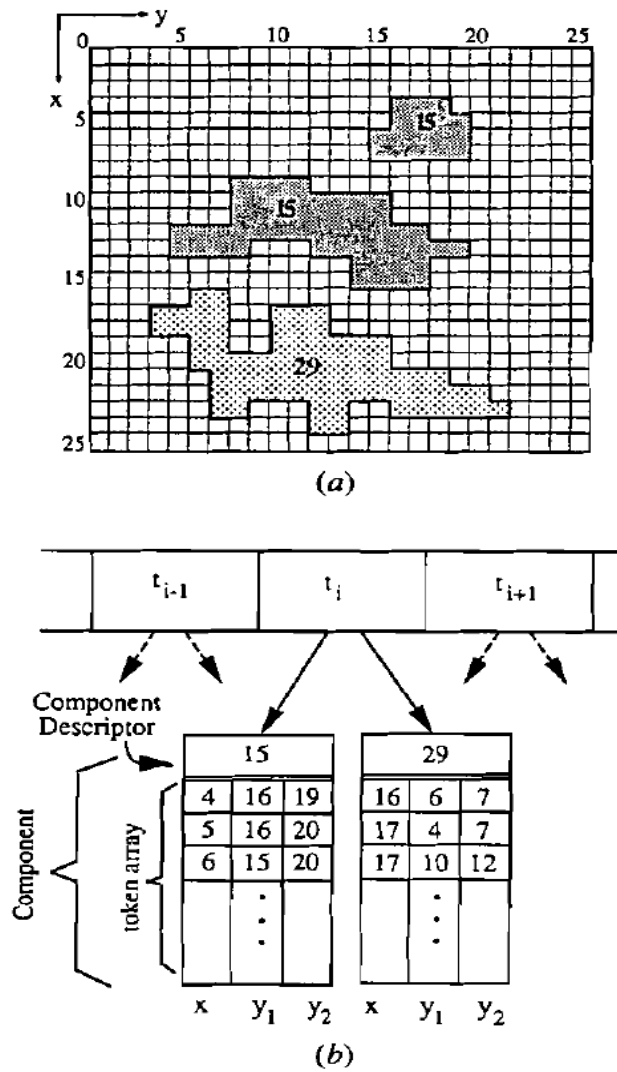


Figure II. 12. Spatial changes at time t_i displayed as a simplified map(a) and the corresponding event components(b). (Peuquet, 1995)

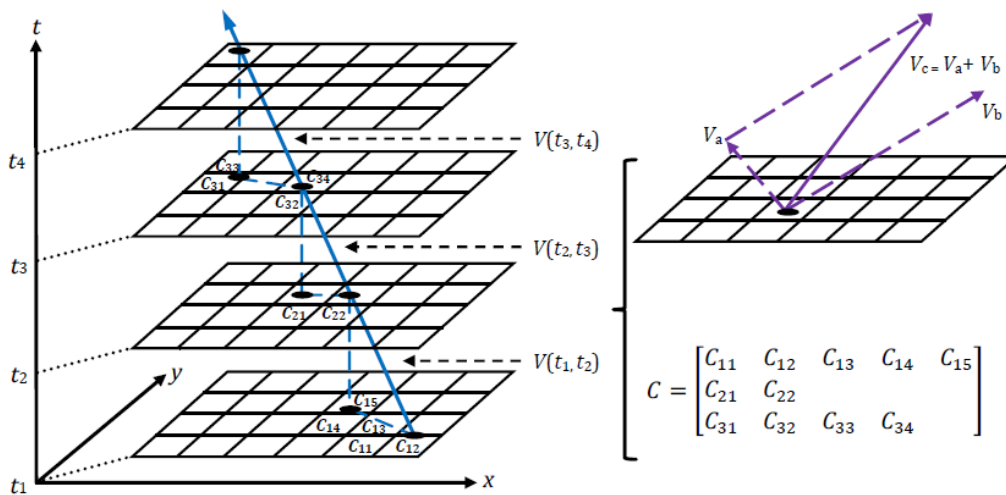


Figure II. 13. Intersecting trip vectors with grids (Liu 2015)

II.4.5 Novel approaches

For instance, Hornsby & Egenhofer (2000) introduced a model for moving objects and that takes into account multiple granularities in time as well as multiples scales in space. That concept of granularity is likely to offer many modelling opportunities and challenges for the development of formal models at different levels of abstractions (Stell & Claramunt 2011). According to a desired granularity, the developed model should be developed according to different levels of a given lifeline. Praing & Schneider (2007) provided a data model to combine historical and future trajectories, this providing a potential support to merge and mixt the data types of moving objects.

Besides trajectory modelling, other researches concentrate on optimizing the operations and queries for trajectory data. For example, Pfoser et al. (2000) proposed a data index oriented to the physical representation and manipulation of trajectory data. Bogorny et al. offer a semantic query language for trajectory data. The semantic language is based on the stops and moves model introduced by Alvares & Spaccapietra (2007) and whose manipulation capabilities is to identify patterns and outliers. An extension for Oracle 11g has been developed to include new spatial temporal objects and operations (Zhao & Jin, 2011). Several algorithms have been developed and introduced in Oracle 11g, as well as the closing exam will give a

brief but hopefully useful summary of the average students' knowledge and expertise. However, and to the best of our knowledge, this modelling extension does not take into account all the semantics of the represented trajectories and their relationships with the surrounding spaces.

II.5 SUMMARY AND DISCUSSION

In order to provide a global view of the different modelling approaches so far identified and discussed in the previous section let us define a series of criteria whose objective is to categorize and characterize them. Without loss of generality we identify some criteria that help to perform this objective, considering of course that they are by no mean absolutely complete and definitive, our objective being to identify commonalities and differences. The criteria considered are as follows:

- **User to System Mediation:** This criterion represents the capability of a given modelling approach to act as mediation between user requirements and the system design level. In other words this criterion characterizes high-level modelling approaches close to the user and the ones close to the application level.
- **Ontology:** Ontologies are knowledge repositories that enable meaningful inter-agent communication in semantic web frameworks (Spaccapietra & Parent, 2009). The objective of this criterion is to reflect to which degree the modelling approach considered takes into account domain-based or generic ontologies.
- **Event:** An event can be defined as a set of related changes leading to a new status (Claramunt et al., 1998), although other definitions have been suggested by other authors. This objective of this event criterion is to qualify modelling approaches that take into account this modelling abstraction as a primitive modelling object.
- **Process:** A process is a concept developed by scientists to understand and relate changes occurring in nature (e.g., soil erosion, orogenic, growing processes)

(Claramunt et al., 1998). It is an intricate mix of facts (status of entities) and transformation mechanisms (ordered changes) that must be considered to structure knowledge about evolution, build models and forecast future situations. Processes can also be considered as components of event. This criterion reflects to which degree a given modelling approach takes into account the concept of process as a primitive modelling object.

- **Hierarchy:** A hierarchy is an arrangement of items (objects, names, values, categories, etc.) in which the modelling items are represented as being “above”, “below”, or “at the same level as” one another (Qian & Yi, 2016). This criterion reflects to which degree this notion of hierarchy is taken into account by the considered modelling approach.
- **Semantic:** The objective of a semantic modelling approach is to search for the conventions and definitions that closely capture the meaning of the user’s ideas (Rishe, 1992). Bishr (1998) suggested that semantics can be defined as the relationships between a computer representation and the corresponding real world context.
- **Graph:** Graph-based approaches for modelling spatio-temporal data provide structural and logical support to relate things, processes and events in a sort of integrated modelling framework. This criterion reflects to which degree a given modelling approach integrates the concept of graph as a structural modelling framework.
- **Trajectory:** A trajectory is represented as a user defined abstraction of the evolution of an object that is moving in space during a given time interval (Spaccapietra et al., 2008). This criterion reflects whether a given modelling approach has been used for trajectory data modelling.

All criteria are evaluated using some general values that characterize if a given modelling approach is compliant with a given criteria (+), possibly (x) or none (-).

Table II. 1. Overview of the spatio-temporal modelling approaches

Modelling Approach	User to System Mediation	Event	Ontology	Hierarchy	Semantic	Process	Graph	Trajectory
Conceptual	+	x	+	+	+	x	-	+
Event-based	x	+	+	+	+	-	-	+
Process-based	+	x	-	x	+	+	-	x
Logical	-	-	-	-	-	-	+	+

Overall our study shows that the spatio-temporal modelling approaches developed so far offer comprehensive and practical methods to represent the behaviour of many trajectory-based applications (Claramunt & Stewart 2015). The extent and principles of the different methods developed and applied to the representation of trajectories in geographical spaces is very large.

When considering the amount of previous research works oriented to the database representation of human trajectories, several travel patterns might be potentially discovered. However, it appears from that study of related work that there is still a lack of specific and practical means to discover travel routines of people in the city that can be easily performed and applied on top of current geographical information systems. Therefore, it appears significant to develop a conceptual and logical framework that will favour further analysis of travel displacements in urban environments. In order to accomplish that, the objective of the research developed in this research is to establish a travel displacement model based on trajectory data.

Since the modelling of human travel displacements is based on trajectory data, a trajectory data based model is to be designed. A large amount of various and heterogeneous trajectory data makes it possible for researchers to dig and analyse different and large kinds of information coming from different fields and expertise. Over the past few years' research in trajectory data modelling has made great progress. However, most of these approaches have some specific advantages as well

as limitations. Given that and as shown previously, the conceptual data model is a useful solution to bring the gap between the analysis and user comprehension, as well that it is a practicably approach to manipulate spatio-temporal data, the framework that will be developed is based on a conceptual spatio-temporal database model.



Travel Displacement Modelling

III.1	MODELLING RESEARCH OBJECTIVE	36
III.2	MODELLING PRINCIPLES	36
III.3	TRAVEL DISPLACEMENT MODEL.....	41
III.3.1	TRAJECTORY MODULE	43
III.3.2	TRANSPORTATION NETWORK MODULE	46
III.3.3	TRAVEL PATTERN MODULE.....	48
III.3.4	RELATIONSHIPS BETWEEN MODULES	54
III.4	DISCUSSION.....	56

III.1 MODELLING RESEARCH OBJECTIVE

Our modelling research objective is to design a generic semantic and spatio-temporal model for human displacements and whose aim is to provide a computational framework that will allow further analysis. The objective of the semantic part of the model is to abstract the objects identified from the real world, that present an interest from an application point of view, and project them into the “information world”. This implies to make a difference between the attributes of the objects of interest identified and the spatial and temporal dimensions. The semantic travel displacement model will act as a foundation to establish a computational representation. The modelling of human displacement is based on the notion of trajectory. The assumption supporting this approach is that an integrated model should facilitate to have a better understanding on the travel displacements, and thus favouring further analysis of citizens’ travel patterns in the city.

III.2 MODELLING PRINCIPLES

As mentioned by Parent (1998), geographical applications are difficult to model due to the inherent particularity and complexity of the spatial and temporal components that emerge from the underlying phenomena to represent. The appropriate modelling of geographical applications requires several advanced abstractions and facilities such as:

- objects that materialize the complex information structure of real-world entities represented including their spatial, temporal and semantic relations that associate them to other modelling abstractions (e.g., composition/aggregation links, generalization links, relationships), as for instance supported by current object-oriented or extended relational models;
- Complementary geometrical features, to support for instance both discrete and continuous views of space, representations at different scale/precision and that allow multiple user points of view;

- Explicit relationships that denote structural links between the entities and abstractions represented: spatial metrics and qualitative relations (e.g., topological (Egenhofer 1991, 1994), directional (Freksa 1993), temporal metrics and algebra (Allen, 1995), and semantic relations;
- Additional modelling abstractions to support multiple temporal granularities and spatial scales (Camossi, 2006).

The object based displacement conceptual model has been developed on top of the modelling method MADS (Modelling of Application Data with Spatio-temporal features) developed by Parent et al. (2006). MADS adopts the object-relationship paradigm, which includes the features of the ODMG (Object Database Management Group) standard model for object-oriented systems (Cattell 97). MADS in fact provides a series of modelling abstractions and sufficient level of flexibility to take into account most of the specific properties of a geographical and temporal application. Overall, the main MADS-oriented modelling concepts used by our approach are as follows:

- An **object** represents a real-world entity. An object type describes a set of objects with similar structure and pattern.
- A **relationship** denotes a link between two or more objects, where each object plays a given role. A relationship describes a set of links with similar characteristics (i.e., linking objects with similar types, roles and properties). The **cardinality** of a relationship denotes the number of occurrences in one object which are associated to the number of occurrences in another. There are three degree of relationship known as one-to-one, one-to-many and many-to-many.
- An **attribute** represents a real-world property; both object types and relationship types may have attributes.
- A **generalization link** is a bottom-up link where several lower level object types, denoted as subtypes, combine to form a higher level object type denoted as super type. **Specialization** is opposite abstraction to Generalization. It is a

top-down approach in which one higher level object type is broken down into two lower level object types.

- An **aggregation link** is a special directed n-ary relationship whose semantics expresses that objects of a type, called composite objects, represent aggregates of objects of another type, called component objects. As for generic relationships, cardinalities are attached to the composite and component roles.

As our travel displacement model should concern data types covering semantics, spatial and temporal aspects, and our approach applies the specialized spatial, temporal and semantic objects and relationships as defined by MADS.

A **spatial** (respectively **temporal**) object type is an object type that holds spatial (respectively temporal) information pertaining to the object itself as a whole (as opposed to pertaining to one of its components or characteristics, or to a relationship between the object and other objects). Similarly, a spatio-temporal object type holds spatio-temporal information.

A **spatial object type** represents a set of spatial objects having the same characteristics. In particular, a spatial object type represents a type of entity whose spatial properties must be represented by defining its shape (e.g., point, polyline, or polygon,) and location. The spatiality of an object is described by a spatial abstract type.

A **Point** is a 0-dimensional geometry denoting a punctual location in space. A **Point** can be used to represent, for instance, a transportation node in an urban network.

A **Polyline** is a 1- dimensional geometry denoting a set of connected points possibly defined by one or more linear (in)equations. It is a connected sequence of line segments. This includes straight and curve lines, polylines, as well as open and closed polylines. A **polyline** can be used to represent, for instance, a transportation segment in an urban network.















An **Oriented Polyline** is a polyline whose sequence of line segments is oriented. An **Oriented Polyline** can be used for instance to represent a trajectory.

A **Region** is a 2-dimensional geometry denoting a set of connected points that lie inside a boundary formed by one or more disjoint closed polylines. A **Region** can be used to represent, for instance, a district in an urban area.

MADS also allows to describe the temporal characteristics of real world phenomena. An **Instant** denotes a single point in time. It can be used to represent, for instance, the time at which an individual arrived at a given Point of Interest (POI). An **Interval** denotes a set of successive instants enclosed between two instants. An Interval can be used to represent, for instance, the lifespan of a trajectory performed by a moving object, from its origin time to its destination time.

Spatial, temporal, and spatio-temporal relationship types hold spatial and/or temporal information pertaining to the relationship as a whole, exactly as for an object type. These relationships can be constrained by space and/or time predicates. The 9-intersection model (Egenhofer 1991, 19994) or the 9-intersection model (Clementini 1994, 1996) is commonly used for defining topological predicates. The Clementini's model is the one used in MADS, as well as in the ISO TC211 and the SQL/MM standards. An example of topological predicates is *Topological Adjacent* linking some geometry objects to denote the *touch* topological relationship. This kind of relationship can act as a constraint to force a transportation node to be associated to a corresponding transportation segment when acting as the termination of that transportation segment.

Table III. 1. MADS spatial and temporal object types and icons

Object & Relation Type	Category	Icon
Object	Generalized Object	
Generalized Spatial Object	Spatial Object Type	
Region	Spatial Object Type	
Line	Spatial Object Type	
Oriented Polyline	Spatial Object Type	
Point	Spatial Object Type	
Generalized Time	Time Object Type	
Instant	Time Object Type	
Interval	Time Object Type	
Relation	Generalized Relation	
Is-A Link	Generalized Link	
Aggregation	Generalized Relation	
Topological Generic Relation	Spatial Relation	
Topological Adjacency	Spatial Relation	

III.3 TRAVEL DISPLACEMENT MODEL

The conceptual database schema of human travel displacements is made of several schemas closely connected (so-called modules hereafter). Each one of the three modules identified plays a specific role regarding the analysis of travel displacements. The design of the model is inspired by the trajectory ontology proposed by Yan et al. (2008) and the construction of the model is designed according to the principles of the MADS model developed by Spaccapietra & Parent (2006).

The conceptual schema comprehends three complementary parts designated as modules:

- The **Trajectory module** is an important component of a human displacement. The Trajectory module abstracts the movements of human beings in the city. Since a large part of people outdoor movements generally happen along the transportation network, we first postulate that the concept of trajectory is closely connected to a network. Figure III.1 shows an example of trajectory decomposed in travel segments qualified by different transportation modes and an overall activity which is ‘going to work’.

- The **Transportation Network module** represents all kinds of network that facilitate human movements inside the city.

- Next, and as to explore human travel displacements, the different factors that generate moving patterns are classified and organized in a module denoted as **Travel Patterns module**.

These three modules (Trajectory module, Transportation Network module, and Travel Patterns module) are not independent as parts of the entities identified are also related by relationships. This is the union of these tree modules that gives the general Travel Displacement schema illustrated in Figure III.2.

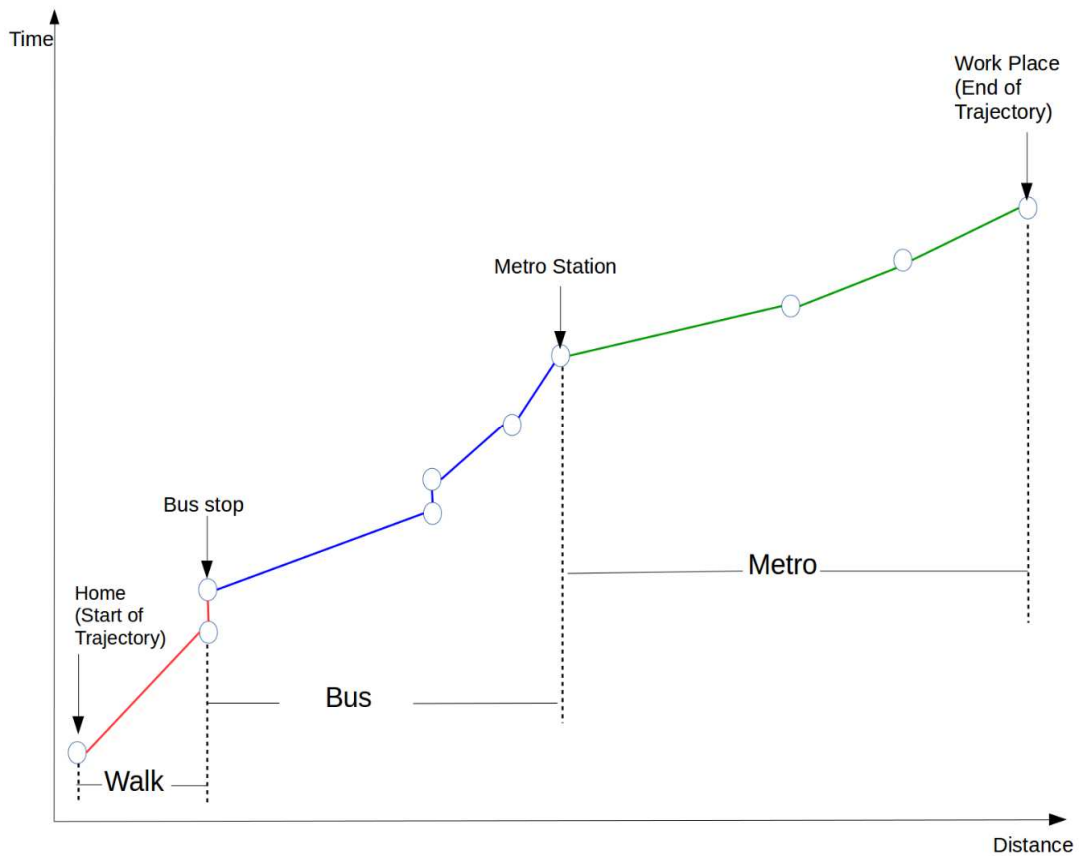


Figure III. 1. Trajectory example

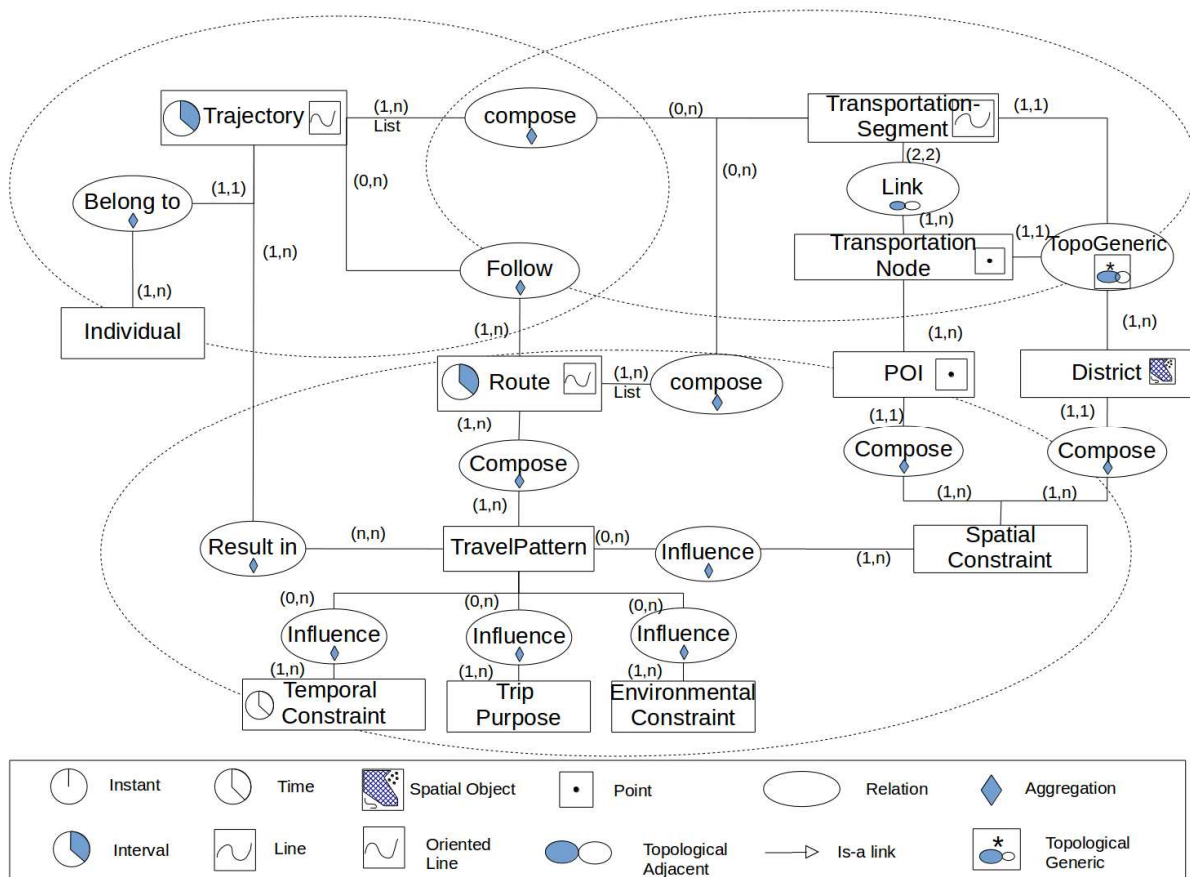


Figure III. 2. General Schema of the Travel Displacement Model

III.3.1 Trajectory Module

The Trajectory Module is based on the trajectory model proposed by Spaccapietra and Parent (2008), in which the concept of trajectory is based on a sequence of stops and displacements as also suggested by early concepts of Time Geography (Hägerstrand, 1970). As we consider people movements in the city that take place along a transportation network, a trajectory is indivisible of a transportation network (Figure III.3). The object Trajectory models a human being movement. Humans are the main actors behind the issue of travel displacement. While early studies in human geography mostly considered transportation displacement at the aggregated level, a conceptual shift appeared in the early 70s with the emergence of the theory of Time Geography (Hägerstrand, 1970). While studying human migrations in Sweden, Hägerstrand developed a conceptual

foundation where human travel displacements are modelled at the disaggregated level.

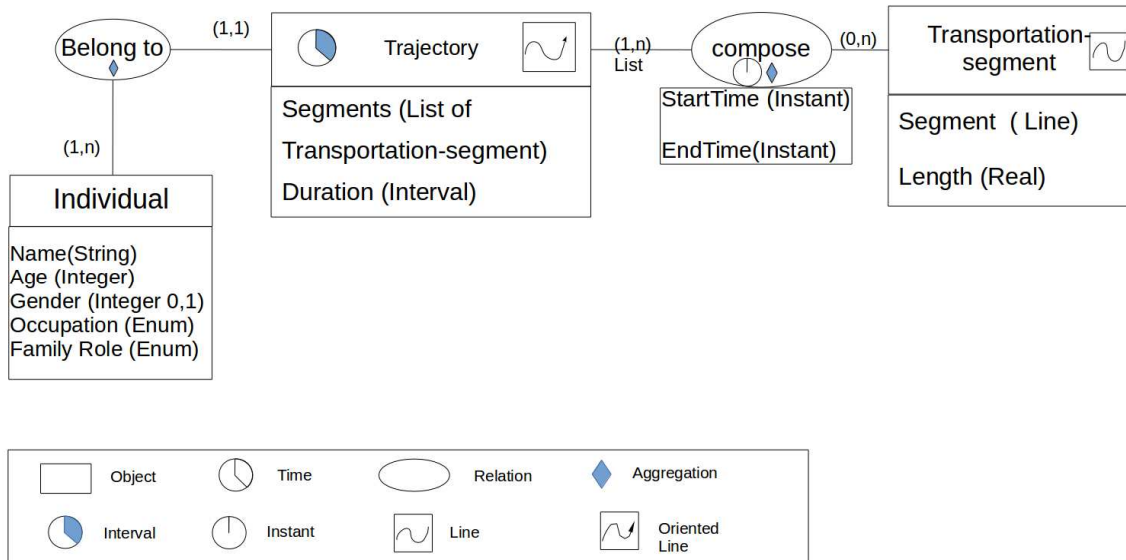


Figure III. 3. Trajectory Module

Following some of the principles identified by the Time Geography, a human hereafter denoted as an *Individual* is an object type representing one of the owners of a given *Trajectory*. A *Trajectory* is spatially and temporally valued.

Definition 1: Individual

An *Individual* is considered as a human being who travels and performs some activities in the city along an urban transportation network. An *Individual* is modelled by several attributes that cover the identity and some social characteristics including *Name* valued as String, *Age* valued as Integer, *Gender* valued as String and two possible values (i.e., *male* or *female*), *Occupation* valued as Enumeration (i.e., an enumeration of possible current occupations), and *FamilyRole* valued as Enumeration (e.g., *mother*, *father*, *child*).

Since the *Individual* is the one who perform the movement, the *Individual* is considered as the owner of the *Trajectory* and the travel displacement. The relationship between *Individual* and *Trajectory* is *BelongTo*. Considering that one *Individual* is very likely to generate several trajectories at different times, the

cardinality between *Individual* and *BelongTo* is (1, n) as a trajectory cannot exist without a moving object, in this case an *Individual*. Without loss of generality, we assume that a given *Trajectory* is generated by one *Individual* only, and the cardinality from *Trajectory* to *BelongTo* is (1,1).

Definition 2 Trajectory

Trajectories imply users' outdoor movements, including the transportation mode utilized. They are for example generated from GPS trajectories, each one represented by a sequence of time-stamped points (instant, latitude, longitude and altitude). A *Trajectory* is designed as an ordered sequence of *TransportationSegments* and is modelled as $Trajectory_i = [TS_1, TS_2, \dots, TS_n]$ where TS_1, TS_2, \dots, TS_n denote a sequence of *TransportationSegments*. Note that a given *TransportationSegment* can be travelled by one to many *Individuals*, which represents the basic modelling unit of human travel. A *Trajectory* has implicitly a starting node and an ending node given by the start node and ending node of its first and last *TransportationSegments*, respectively.

The relationship between *TransportationSegment* and *Trajectory* is *Compose*, defined as an aggregation of type List and it contains two temporal attributes, *StartTime* and *EndTime*, both of type Instant. These two attributes describe the start and end times of the trajectory segment. Indeed, this means that a given *Trajectory* is a sequence of one-to-many *TransportationSegments*. This denotes the fact that *TransportationSegments* are static spatial objects. On the contrary the trajectory which reflects the movement of moving object along the transportation network is dynamic. The connection between *TransportationSegments* and *Trajectory* is actually implicitly related to a time interval which denotes the time spent by this trajectory on that *TransportationSegments*. Therefore, with the *TransportationSegments* and the start and end time of each related time interval, the trajectory of a moving *Individual* is clearly defined. Furthermore, taking into consideration that there might be several trajectories going through the same *TransportationSegment* and at various dates and times, a database instance is very likely to contain several instances of *TransportationSegment*. For instance, let us assume a trajectory T1 of a girl leisurely walking in Changan Street on March 20, 2016, between 6:15pm and 6:40pm, and a *Trajectory* T2 of a man driving a car on Changan Street, the same day, March 20, 2016, between 6:20pm and 6:23pm. These two trajectories go through the same transportation segment, but at different times,

this generating two object instances *TransportationSegment1* and *TransportationSegment2*. Therefore, the cardinality between *Trajectory* and *Compose* is (1, n). On the other hand, as theoretically it might be possible that a *TransportationSegment* is not taken by any Individual recorded, as well as it might be passed by multiple *Trajectories*. So the object *TransportationSegment* is not necessarily assigned to one or more *Trajectory*. The cardinality from *TransportationSegment* to the relationship *Compose* is (0, n).

A *Trajectory* also contains a temporal attribute, *Duration*, valued as an interval and derived from the time instant of the starting node of the first *TransportationSegment* TR_1 , and the time instant of the ending node of the last *TransportationSegment* of that *Trajectory* $Trajectory_i$. This particular *Duration* records the total time it takes for an Individual to pass through a sequence of *TransportationSegments* denoting a *Trajectory*.

III.3.2 Transportation Network Module

The Transportation Network Module models the extension of the transportation network of a given city. For the objectives of our application it contains the road and railway networks. Along the road network, people walk, drive, takes taxis and buses. The railway network includes tramway and metro lines. In the Transportation Network Model, instead of defining every type of transportation network separately, we generalize all kinds of Transportation as a whole. Accordingly, the transportation network is summarized to all means of transportation, while this information is recorded at the attribute level, but redundancy is avoided. However, this modelling method also supports further extensions to take into account additional transportation means with some minor adaptations.

Definition 3: TransportationNode

A *TransportationNode* is a point in the Transportation network, that is, a generalisation of either a transportation line intersection or a *POI*. A transportation line intersection denotes the junction between several *TransportationSegment* (at least two). A *POI* is a point of interest which might approximates the location of a place of interest and landmark in the network, a destination or a stop of a trajectory (e.g., home in figure III.1).

The object type *TransportationNode* has a spatial attribute *Node* defined as a geometry of type Point which represents the location of that *TransportationNode* in the Transportation Network. A Boolean attribute *IsStop* differentiates a *TransportationNode* that models a transportation intersection (i.e., valued as 1) from a *POI* (i.e., valued by the Null value) in the urban network. *IsRoad*, *isMetro* and *isRailway* are the attributes that differentiate the transportation network type the node belongs to. There is also the case that one *TransportationNode* can denote a cross of multiple *TransportationSegments* and that of different types of transportation including *Road*, *Metro* or/and *Railway*. Therefore, we value the *TransportationNode* with these three attributes to record the type or multi-type of transportation of that node.

Definition 4: TransportationSegment

A *TransportationSegment* models a link between two *TransportationNodes* in the transportation network. The spatial attribute of *TransportationSegment* is a finite polyline along the Transportation Network, ended by two *TransportationNode*. The length of the geometry polyline is also recorded as an attribute. The attribute *TranspoType* defines the different types of Transportation Network, such as *Road*, *Metro* and *Railway*. An attribute denotes the type of *TransportationSegment* as a *TransportationSegment* can only belongs to only one type of transportation. When even more than one *TransportationSegments* are overlapping, we consider that they are different objects.

The relationship between a *TransportationSegment* and a *TransportationNode* is defined as a topological relationship *Touch* (Figure III.4) between a polyline and a point. A given *TransportationNode* is in relationship with one-to-many *TransportationSegments* while one given *TransportationSegment* should have two *TransportationNodes* that denote the spatial origin and destination of that segment, respectively. Therefore, the cardinality from *TransportationSegment* to *Touch* is (2, 2), and from *TransportationNode* to *Touch* is (1, n).

Apart from the modelling constructs identified in the conceptual database schema presented in Figure III.1, an additional spatio-temporal constraint is taken into account in order to reflect the complete semantics of the application. This constraint is as follows:

Constraint 1

All successive *TransportationSegments* that compose a given *Route* must be connected, that is, either the starting *TransportationNode* or the ending *TransportationNode* of that *TransportationSegment* is either the starting or ending *TransportationNode* of another *TransportationSegment* of that *Route*.

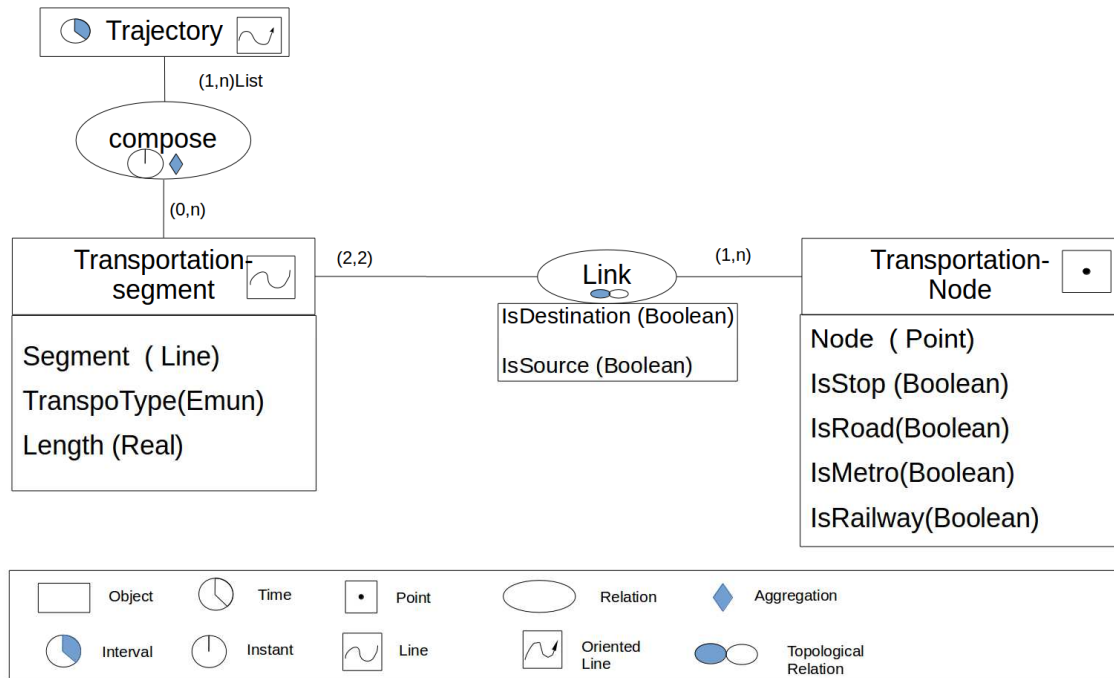


Figure III. 4. Transportation Module

III.3.3 Travel Pattern Module

TravelPatterns are considered as elements or components of the modelling of travel displacements (Figure III.5).

Definition 5: Route

A *Route* represents a path chosen by one-to-many *Individuals*. Let us consider an *Individual_i* that travels along a transportation network and that generates a *Trajectory*. The *Route* is generalized from parts of *Trajectories* that repeat routinely. Every *Route* starts at a *Node* and ends at a *Node*. All the *TransportationSegments* and *TransportationNodes* passed through by are implicitly temporally ordered. More

formally a *Route* is modelled as a sequence of *TransportationSegments*. In order to represent the spatiality of a *Route*, we introduce the concept of *Path* defined as a sequence of *TransportationSegments*.

$$Path_i = [TransportationSegment_1, TransportationSegment_2, \dots, TransportationSegment_n];$$

In order to represent routine movements, we introduce a temporal regular interval that represents repetitive temporal patterns not pointing to a specific time in the calendar but a repetitive one.

$$TemporalCycle = \langle Regularity, Regular\ Interval \rangle$$

An example of *TemporalCycle* is *TimeSegment* = $\langle workdays, (8:00pm, 9:00pm) \rangle$.

Then a route is defined as follows:

$$Route_i = \langle Path_i, TemporalCycle_i, TranspoType \rangle$$

An instance of *Route* could be as follows:

$$[\langle TransportationSegment143, \langle workdays, (08:00, 09:00) \rangle, Bus \rangle, \langle TransportationSegment659, \langle workdays, (08:00, 09:00) \rangle, Subway \rangle, \langle TransportationSegment658, \langle workdays, (08:00, 09:00) \rangle, Walk \rangle]$$

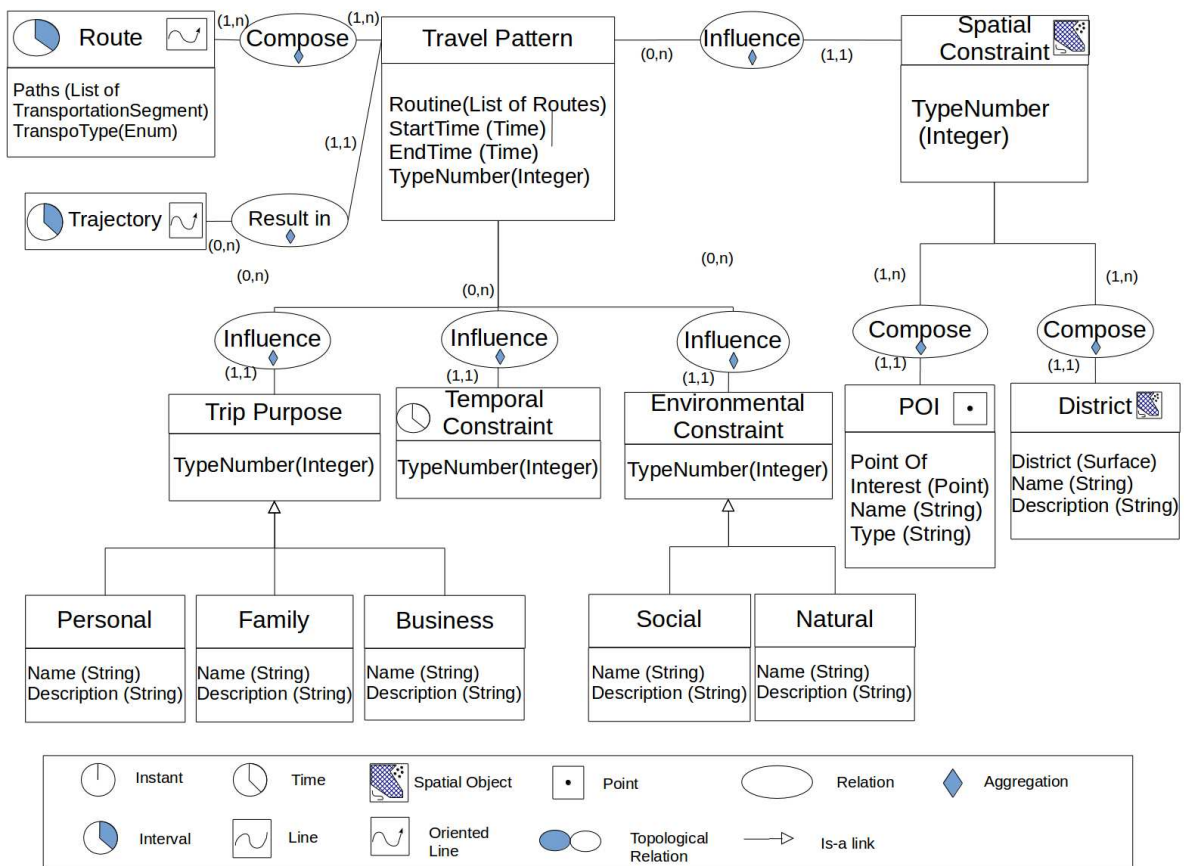


Figure III. 5. Travel Pattern Module

In other words, while an *Individual* can perform a *Trajectory* only once, he/she can follow the same *Path* many times, as well as he/she can make the same *Route* on a regular basis.

Definition 6: TravelPattern

A *Route* from a Node N_1 to a Node N_m can be shared by several *Individuals* $\{i_1, \dots, i_n\}$. A given *Route* can be caused by different or similar Constraints (*TripPurpose*, *TemporalConstraint*...). *Routes* generate *TravelPatterns* when performed on either regular or non-regular frequencies.

Trajectories contain different *Routes* whose analysis reflects different *TravelPatterns*. While *Trajectories* and *Routes* are explicitly and precisely represented in space and time, *TravelPatterns* are derived from *Routes* according to some frequency constraints.

TravelPattern = $\langle \text{Routine}:[\langle \text{Route}_1 \rangle, \langle \text{Route}_2 \rangle, \dots, \langle \text{Route}_n \rangle], \text{TemporalCycle}, \text{Frequency}, \text{TypeNumber} \rangle$

A *TravelPattern* should reflect the different factors that influence an *Individual's* travel decision. When selecting a *TravelPattern*, several factors influence a human decision. For example, when it's raining in the morning on a workday, an *Individual* might choose to take a car to go to work, while he/her might walk and take a bus on a regular basis. The attributes of a *TravelPattern* contain the individual identifier, and the different constraints that influence this *TravelPattern*. These constraints includes environment constraints (weather, scenery, road condition...), temporal constraints (morning, afternoon, weekdays...), spatial constraints (district, neighbourhood, POI...) and trip purpose (shopping, going to work...).

Overall, a *TravelPattern* is mainly generated and influenced by these external and internal constraints. These represent the basic displacement foundations of our modelling approach.

Travel displacements vary according to *Individuals* as the range of possible actions and activities is extremely large and reflects different socio-cultural backgrounds. Although our objective is not to develop an exhaustive model that will reflect all possible parameters that determine these patterns – this being surely unrealistic – we nevertheless decided to select a few objectives and individual attributes usually represented in social science studies such as gender, age, occupation, salary and role in the family. According to Hågerstrand (1970), human behaviour is often constrained by some social obligations. Therefore, it is reasonable to believe that these basic personal profiles can be used to depict some human categories that will reflect different backgrounds and then human travel displacement.

For example, by considering different economic conditions and occupations of a given family, travel modes might vary as a student is likely to take different travel decisions when compared to a full time adult employee. Moreover, since an *Individual* generally perform multiple travels and *Trajectories* over a long period of time, an *Individual* itself is likely to influence several *TravelPatterns* during that period. And an *Individual* can have more than one *TravelPattern*. Nevertheless, since the *Individual* is also connected with *Trajectory*, we assume that it is the *Trajectory* that influences the *TravelPattern* directly. For instance, a given individual can have travel patterns of kind "going to work" on the morning and late

afternoon of weekdays, and travel patterns of kind "going to restaurant" in the evening, which are reflected by *Trajectories*. Thus, we link *TravelPattern* to *Trajectory* instead of *Individual*. In this case each *Trajectory* would be characterized by its own travel patterns.

There are many factors which can influence *Individual's* travel decision that forms *TravelPattern* eventually. We generally classify these factors into five categories, which include *TemporalConstraint*, *TripPurpose*, *EnvironmentConstraint*, *SpatialConstraint*. Indeed, individual social characteristics as defined in the model also play an important role in the way a given individual behaves in the city. Overall, an *Individual* is likely to own several *Trajectories* while each *Trajectory* reflects some *TravelPattern*. These different types of constraint show different factors that can affect a travel pattern. In the Travel Pattern Module, some categories are spatial constraints and are denoted as *SpatialConstraint* and can include *POIs* and *District*. *POIs* might be for example a place where a given Individual plan to pass through when moving in the city (e.g., a supermarket), this place playing the role of a spatial constraint. Another example of spatial constraint might be given by the *District* where he/she lives or well known to him/her, that district being privileged when he/she moves in the city.

Regarding the impact of a *TemporalConstraint* over a *TravelPattern*, this represents different times such a day/night, rush hour/or not, weekday/weekend, holidays/or not. Indeed, the time of travel is likely to affect a *TravelPattern*. For example, if a girl travels walking and alone at night, she is likely to prefer a safer street. *TemporalConstraint* is made of different temporal granules that reflect the intervals mentioned above.

A *TripPurpose* is used to represent the purpose of travel, including the things planned to be done by the *Individual* through traveling. For example, one plans to go to work at 8:00 am. In this case, "go to work" is the *TripPurpose* in the *TravelPattern* module. The activities of a human can be classified into sub-categories of *PersonalActivity*, *FamilyActivity* and *BusinessActivity*. Generally, different types of activities often result in different movement habits. For a *BusinessActivity*, minimizing the time travel is often an important criterion to apply, while for a *FamilyActivity* this might not be the case.

The environmental context might also impact people's mind and decisions when travelling, this generating *TravelPatterns*. We classify *EnvironmentalConstraints* into two categories, that is, the *Social- Environment* and

the *NaturalEnvironment*. The *SocialEnvironment* models some happenings close to the travel environment, while the *NaturalEnvironment* includes some physical characteristics such as weather, landscape and so on. Overall, *TemporalConstraints*, *EnvironmentalConstraints* and *TripPurpose* are qualified by some specific attributes valued using some Tuple constructs and specific attributes. In the MADS model, an is-a link is a type of generalization link relating subtypes to supertypes, and supertypes attributes are inherited by their subtypes. Thus, the relationship between the objects like Environment, Social and Natural is assigned ISA (is-a link).

Last but not least, *SpatialConstraint* models the spatial properties that can impact the *TravelPattern* of an *Individual*. As for the constraints previously introduced, *SpatialConstraint* represents a category of factors influencing travel decisions. A *SpatialConstraint* can include points of interest (*POI*), districts where people choose to travel and additional *SpatialConstraints*. Without loss of generality, *POIs* and *District* are taken as examples to illustrate a *Spatial-Constraint*. A *POI* is a location of interest where someone finds it useful or interesting and worth a visit. A *POI* is spatially modelled as a point that denotes the location of this place people are interested in, such as a restaurant or a cultural monument. These places are most likely to become a source or a destination of a *Trajectory*. *POI* attributes may implicitly give some reasons for people to travel to this point of interest. *District* is another type of spatial object that can influence people's travel decisions. A city is usually divided into several districts for administrative reasons. People's movements in the city are partly constrained by *Districts*. *District* is considered as a geometry Surface. These factors that can affect travel displacements are denoted by appropriate spatial, temporal and spatial attributes, and spatial relationships that relate them to *Trajectories*.

Constraint 2

The union of all the Districts forms a partition of the considered region of interest. Each District has no overlaps with any other Districts.

Constraint 3

A *TransportationSegment* can overlap one-to-many *TransportationSegments* whatever the transportation mode.

Constraint 4

A *TransportationNode* derived from a *POI* should only result from an intersection of that *POI* with a *TransportationSegment* of type *Road*, that is, the attribute *isRoad* is true. We made this assumption to make sure all *POIs* are associated to the road network.

III.3.4 Relationships between Modules

Relationships are essential abstractions used to design the Travel Displacement schema. Most often connections between the different modules are materialized by specific relations that associate the different modelling abstractions identified. For instance, the *Follow* relationship between the object type *Route* and the object type *Trajectory* is a key abstraction of the *TravelPattern* module (Figure III.6), such as the temporal relationship *Compose* between *Route* and *TransportationSegment*. Therefore, *Route* is the object type that implicitly connects the three modules. The cardinality between *Trajectory* and the line *Follow* is (0, n). This is due to the fact that a *Trajectory* might not follow a *Route*, while on the contrary a *Route* should at least follow one *Trajectory*. As a *Route* is made of a list of *TransportationSegments*, the cardinality from *Route* to *TransportationSegment* is (1, n), while from *TransportationSegment* to *Route* is (0, n).

A *TravelPattern* is characterized by different factors in the *TravelPattern* module. As for the relationship between the Travel Pattern Module and the Trajectory Module, multiple relationships have been modelled. The relationship between the Trajectory Module and the Travel Pattern Module denotes the indirect relation between *TravelPattern* and *Individual*, and finally represents the fact that every *Trajectory* belongs to an *Individual*. Indeed, the *Individual* abstraction plays a crucial role when modelling travel displacements, and it can be seen as a dominant abstraction when modelling a travel pattern. One *Individual* can own one-to-many *Trajectories* to reflect some travel patterns, his/her travel decisions being the result of different constraints. A *District* is in spatial relationship with the Transportation Network. Considering the possible topological relationship for instance between *TransportationSegments* and *Districts*, it appears that the possible relationship include *Touch*, *Within* and *Intersect*. These relations are summarized in the database schema by the topological relation *TopoGeneric* (Figure III.7). *POIs* also play an important role, as *Trajectories*, *POIs* reflect source, intermediate or destination nodes of a given trajectory. A *POI* is connected by an Equal topological relation

with a *Node*, but with the respective *POIs* and *Node* sets being disjoint. Finally, *POIs* are likely to bring many opportunities for further analysis when the whole model is logically implemented and experimented with real data.

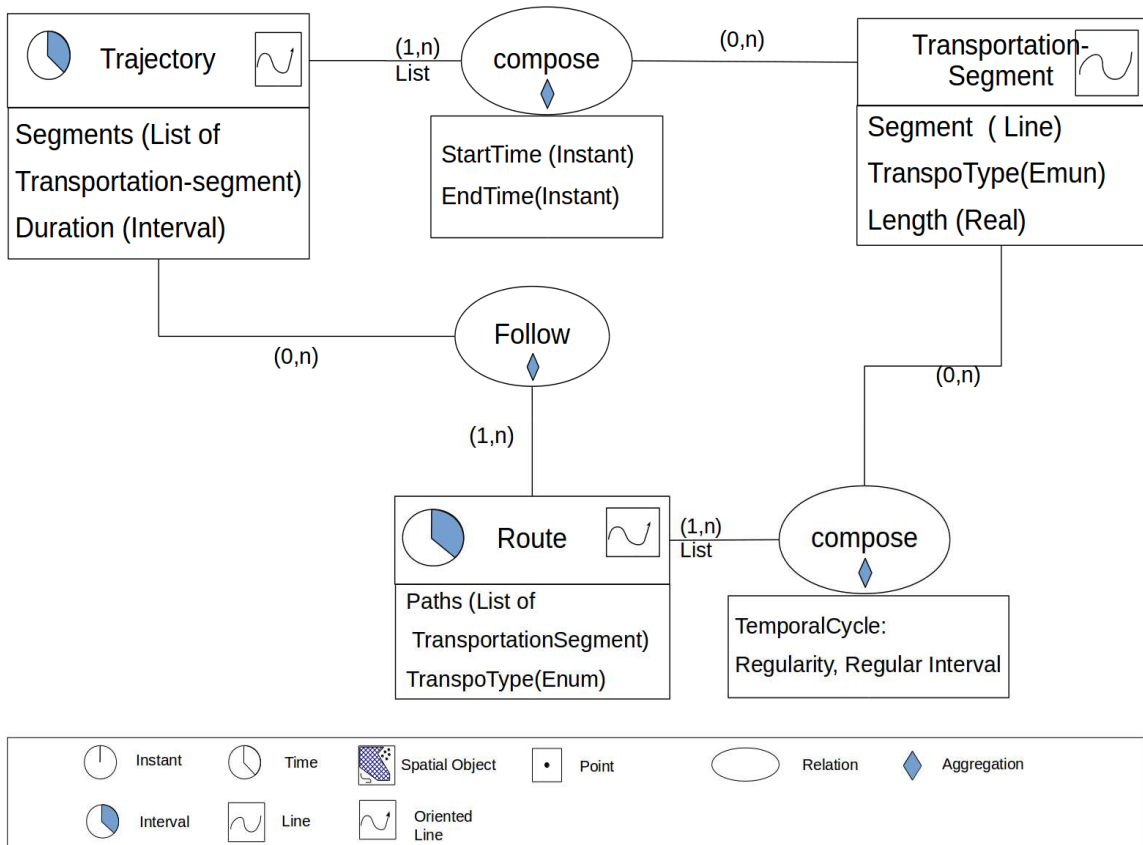


Figure III. 6. The relationship between Trajectory and Route

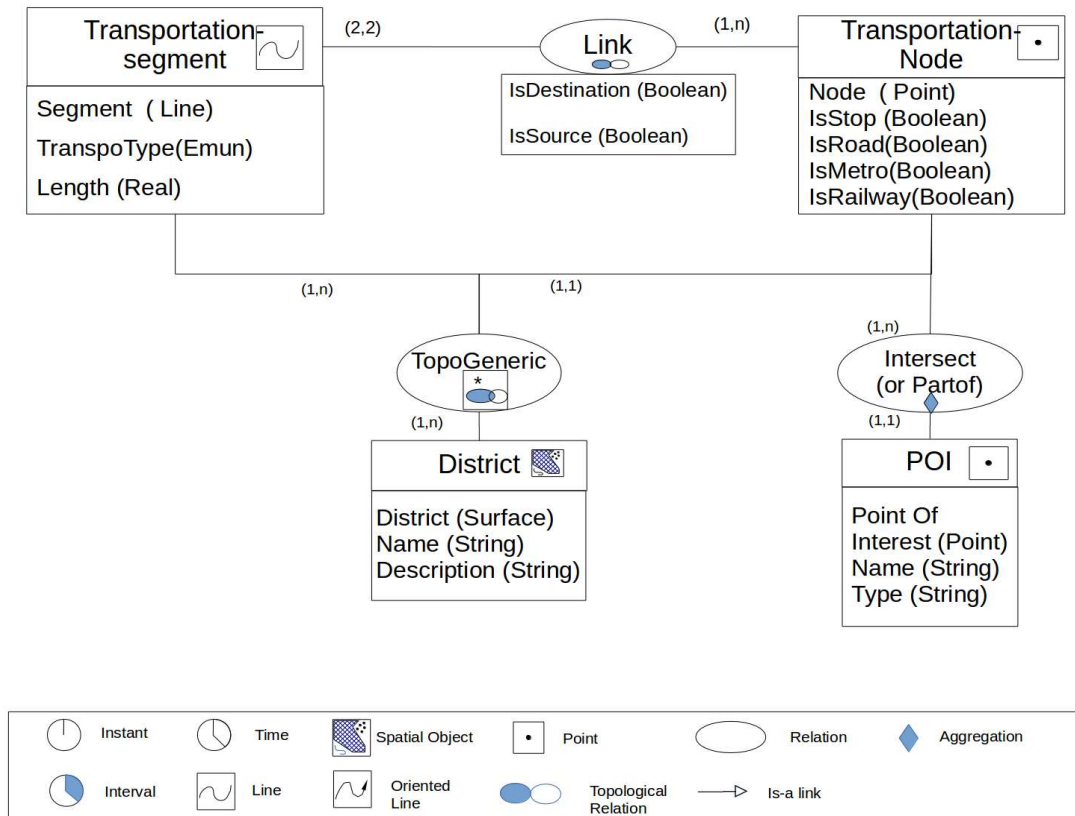


Figure III. 7. The relationship between Spatial Object Types

III.4 DISCUSSION

This chapter introduces the main principles and the different components of our database conceptual approach whose objective is to characterize the main modelling abstractions required to reveal the semantics of our application context, that is, trajectories patterns as revealed by real-world data. The travel displacement data model developed introduces and specifies the different concepts necessary to for an appropriate representation of human trajectories. The model supports the representation of a trajectory object as well as the relationships between individuals, trajectories, and the whole underlying geographical context (e.g., transportation network, routes, and urban layout). Several additional properties factors that play a role when modelling travel displacements are also represented by the conceptual schema. These constraints can be defined at the semantics, spatial and temporal

levels and might be related to the notions of activities or to the underlying geographical environment.

Overall, this model provides a relatively generic conceptual representation of human's travel displacements in a city realm. The different spatial, temporal and semantic elements that compose such displacements are considered and specified, as well as their main attributes and relationships that characterise them. The main interest of this conceptual approach is to provide a complete data structure that represents the complexity of the application considered, and that supports an integration of real world trajectories. Such environment will then be explored in term of its manipulation capabilities in the following chapters. We expect the framework to reveal some travel patterns and specific trends, as well as it will favour further exploration and analysis.



Travel Displacement Database and Analysis

IV.1	DATABASE IMPLEMENTATION OBJECTIVES	59
IV.2	TRAVEL DISPLACEMENT DATABASE	59
IV.3	TRAVEL DISPLACEMENT ANALYSIS.....	65
IV.3.1	DATA ANALYSIS CATEGORIES	65
IV.3.1.1	<i>Semantic level.....</i>	65
IV.3.1.2	<i>Temporal level.....</i>	66
IV.3.1.3	<i>Spatial level.....</i>	66
IV.3.1.4	<i>Spatio-temporal levels.....</i>	67
IV.3.2	DATA ANALYSIS FUNCTIONS AND QUERIES.....	70
IV.4	DISCUSSION.....	83

IV.1 DATABASE IMPLEMENTATION OBJECTIVES

The conceptual model developed for human travel displacements provides a supporting framework to define and represent how people travel and in a city. This modelling framework should support the logical design and implementation of a travel displacement database, and then data manipulation capabilities to explore and apply some data analysis to study travel displacements in the city. Spatio-temporal database systems offer some privileged, popular and efficient solutions to integrate, manage and manipulate different kinds of spatio-temporal and semantics data. This leads us to develop an implementation of a travel displacement database at the logical level, derived from the conceptual model illustrated in the previous sections, this conceived as a spatio-temporal database whose objective is to develop some spatio-temporal data analysis functionalities to analyse travel displacements, patterns and outliers in a given city selected as an illustrative testbed. One specific objective is to discover some travel patterns revealed by people's movement in the city, searching for some regular and routine trajectories. For instance, workday journeys, trajectories patterns over week-ends, regular routes from one neighbourhood to another, temporal and spatial commonalities as well as differences when aggregating trajectories are amongst the sort of people displacement patterns that one might explore. In order to achieve such data analysis objectives, our objective is to implement a human displacement model at the logical level, and then to explore different data query, aggregation and manipulation mechanisms to infer some specific trends at the individual and global levels.

IV.2 TRAVEL DISPLACEMENT DATABASE

The design of a travel displacement model provides a conceptual interpretation and representation of how people travel in a given city. Such design has been developed on top of the MADS conceptual framework, this providing data abstraction and constraints that give a sound support to the model. In particular, data

integrity constraints provide some solid mechanisms to construct the travel displacement model at the logical level, and then to explore and implement some query mechanisms. The logical design and interpretation of the conceptual database model developed in the previous sections are hereafter introduced. For every spatial or spatial temporal object, the spatial reference is defined as Geometry, and also valued by a SRID (Spatial Reference System Identifier) which is a unique value used to unambiguously locate geographical entities according to a given coordinate reference system.

Table IV. 1. TransportationNode Table (Related to the object TransportationNode)

Attribute Name	Data Type	Description
ID	integer	TransportationNode identifier
IsRoad	Boolean	True if that TransportationNode connects one-to-many roads (TransportationSegments with an attribute Type="Road")
IsMetro	Boolean	True if that TransportationNode connects one-to-many metros (TransportationSegments with an attribute Type = "metro")
IsRailway	Boolean	True if the TransportationNode connects one-to-many railways (TransportationSegments with an attribute Type = "railway")
GEOM	geometry(Point,SRID)	Spatial representation as a geometry Point with SRID

Table IV. 2. TransportationSegment Table

Attribute Name	Data Type	Description
Id	Integer	TransportationSegment identifier
Name	character varying	Name of the road or name of the metro line
StartNode	integer	Reference to the TransportationNode table
EndNode	integer	Reference to the TransportationNode table
Length	double precision	Length of the TransportationSegment
Type	character varying	Type of the transportation network (i.e., road, metro, railway)
GEOM	geometry	Spatial representation as a geometry Polyline with SRID

Table IV. 3. Individual Table

Attribute Name	Data Type	Description
ID	integer	Individual identifier
Name	character varying	Individual name
DateofBirth	geometry	Individual date of birth
Profession	character varying	Individual profession

Table IV. 4. TrajectoryMain Table

Attribute Name	Data Type	Description
ID	integer	Trajectory identifier
TableName	character varying	Name of the table that contains the list of the TransportationSegments that the trajectory passed by
IndividualID	integer	Owner of the trajectory
StartTime	datetime	StartTime of the trajectory
EndTime	datetime	EndTime of the trajectory

Table IV. 5. TrajectoryPartTable

Attribute Name	Data Type	Description
ID	integer	Trajectory identifier
TransportationSegmentID (foreign key)	integer	TransportationSegment the trajectory pass through
StartTime	datetime	Time when the individual enters the segment
EndTime	datetime	Time when the individual leaves the segment
IsStart	Boolean	Denote if it is the first TransportationSegment of the trajectory
IsEnd	Boolean	Denote if it is the last TransportationSegment of the trajectory
TranspoType	character varying	Transportation method that the individual used to travel

Table IV. 6. Route Table

Attribute Name	Data Type	Description
ID	integer	Route identifier
IndividualID	integer	The ID of the Individual who performs the Route
Path	integer[]	TransportationSegment array
TranspoType	character varying	Transportation method that the individual used to travel
StartTime	time	The hour of a day the Route starts
EndTime	time	The hour of a day the Route ends
TimeGranularity	Varchar[10]	Time granularity that denotes how often the route is performed (minute, hour, day, week, month)
GranularityCycle	integer	GranularityCycle represents how many intervals of the given Time Granularity denote the cycle.

In the *RouteTable*, the attributes *TimeGranularity* and *GranularityCycle* together denote the *TimeCycle* as represented by the conceptual model. For example, let us consider a route of *TimeGranularity* ‘day’ and a *GranularityCycle*=10, then this means that this route is repeated every 10 days.

Table IV. 7. POI Table

Attribute Name	Data Type	Description
ID	integer	POI identifier
Postcode	character varying(3)	POI postcode
Name	character varying	POI name
Address	character varying	POI address
Function	character varying	Function of the POI (education, shopping mall, supermarket, company...)
GEOM	geometry(Point, SRID)	POI Spatial location

Table IV. 8. District Table

Attribute Name	Data Type	Description
ID	integer	District identifier
Name	character varying	District name
GEOM	geometry(Polygon, SRID)	District spatial representation

Table IV. 9. Structure of the Activity Table

Attribute Name	Data Type	Description
ID	integer	Activity identifier
ActivityType	character varying	Activity type(i.e., personal, family or business)
Name	character varying	Activity name

Table IV. 10. Environment Constraint Table

Attribute Name	Data Type	Description
ID	integer	Environment Constraint identifier
EnvironType	character varying	EnvironType of the environment constraint (weather, road condition or geographic)
Name	character varying	Environment constraint name

Table IV. 11. Weather Constraint Table

Attribute Name	Data Type	Description
ID	integer	Weather identifier
TheDate	date	Date of the weather constraint
Rain	double precision	Precipitation value
Wind	double precision	Wind speed

IV.3 TRAVEL DISPLACEMENT ANALYSIS

IV.3.1 Data Analysis Categories

Travel displacement queries should provide explicit answers to explicit questions related to the exploration of human's movements in the city. Well-designed categories of queries applied to those data sets should organise the different query capabilities and facilitate a search for a better understanding of travel displacements in the city realm. According to Dodge (2008), people's travel patterns can be classified into three categories: **spatial patterns**, **temporal patterns** and **spatio-temporal patterns**. The modelling approach developed by our research also considers the spatial and levels but also the semantics one as indeed the semantic dimension plays a significant role for travel displacement analysis. On top of the travel displacement model suggested, one might expect some traveling analysis to be designed and performed. The queries involved all three dimensions of spatial, temporal and semantic levels.

IV.3.1.1 *Semantic level*

When considering the semantic level, queries usually deal with some generic dimensions that cover the “who”, “what”, “how” of the objects of interest. Within the travel displacement model, the semantic objects and attributes are mainly represented and described in the Travel Pattern Module, which reports the underlying properties associated to travel displacements. In this module, the

semantic factors which denote some of the outcomes of people's travel patterns are specified according to three complementary categories of properties: *Individual Profile*, *Trip Purpose* and *Environmental Constraint*. Usually, semantic queries are very likely to be combined with spatial and temporal constraints.

Although the Travel Pattern Module encapsulates most of the semantic data, or at least the richer part of the semantic data represented, there is still the opportunity to apply a few additional semantic queries on other modules. For instance, taking the object type *Individual*, it is possible to answer the basic question "*Who is that individual?*". Qualifying and recognizing who is a moving individual should help to understand the reasons and regularities under which this individual makes such travel choices in the city. Overall deriving a better knowledge of the individuals that travel in the city should provide a better view of the population dynamics in the city. By further analysing the type of *PointOfInterest* and the types of *Activity* related to these trajectories, the character of a given individual can be roughly studied.

IV.3.1.2 *Temporal level*

The queries applied to the temporal level should help to discover regular or irregular patterns such as when some people start and stop when performing some trajectories and/or activities as well as the time spent (i.e., temporal interval) for performing that trajectory and/or activity. Indeed, temporal queries are essential to find out some routines or irregular travel displacements in the city, which is a valuable potential contribution transportation planning tasks in the city. Some of these temporal queries usually start with "When", "How long", "How often". Such queries mainly should be preferably to the Trajectory Module. For example, "*When a given or some individuals start to travel this morning and when do they get back home?*". Other questions that start with "How often" rather ask about the frequency of some individual's mobility choices. Such queries are particularly relevant for trajectories routine discovery. For example, "*How often a given or some individuals cross the specific road?*"

IV.3.1.3 *Spatial level*

The queries of the spatial level mainly deal with data and attributes related to locations. For instance, spatial entities have an explicit representation of their geometry. Apart from the a basic spatial query "*Find the spatial path of a given trajectory?*", there are other queries of various types that can be performed. These queries might search for some specific geometrical properties such as metrics,

directions or topological relationships. For instance, the query “*Find the longest trajectory performed by a given individual*” is an example based on a metric. Another example related to a topological property is “*Find the trajectories that cross a given district in the city*”, which explores the relationships between trajectories and a given spatial entity.

IV.3.1.4 Spatio-temporal levels

Queries applied to the spatial and temporal levels combine spatial and temporal attributes of the objects are represented by the travel displacement model. They reflect questions related to the “Where and When”. “Where” usually refers to some trajectory spatial properties (e.g., origin and destination locations, path) or even some of the *SpatialConstraints* and *TemporalConstraints* introduced in the previous section. An interesting pattern that might be exhibited by such spatio-temporal queries is that they are likely to implicitly reflect some functional properties in the city. For instance, one might compare the trajectory footprints of a given *Individual* on a daily basis or the *Trajectory* that cross different neighbourhoods on a given day. Further queries might even search for trajectory clusters in space at different times. Some relevant examples can be “*Search for daily differences and commonalities between the trajectories of a given individual over a period of one week?*” and “*Find all the trajectories that passed through a given road on the 6th Sept. 2009*”. The “When” might refer to the start or end time of some Trajectories, or referring to the frequency of some travel displacement’s occurrence. For example, “Where and when a given or some individuals travel during the weekend?”. Travel patterns can be also studied at the aggregate level in order to reflect regular or irregular displacements during a period of time. For example, “*What are the most regular travel paths of today’s travels and how much time did they take?*”

(5) Semantic, spatial and temporal levels

In order to explore travel displacements in the city, queries should combine the semantic, spatial attributes and temporal attributes. The extent of semantic, spatial and temporal queries covers a large range of opportunities, and they can be classified into queries referring to either a single individual or sets of individuals. Queries applied at the individual level are likely to search for irregular or uncommon displacements or specific searches to for instance explore a specific and personal travel planning task. Examples of such queries are as follows.

Who, when, how and where some individuals are moving?

Individual or aggregated displacements can be explored in such a way, which for example can be based on the regions or places an *Individual* usually moves around during different time intervals. For example, if an *Individual* moves around a campus early in the morning and late in the afternoon, then there is a high probability that/she is a student or a teacher. Next, as far as *Individual* with similar characteristics can be identified and classified, one might explore and suggest travel routes and trajectory decisions for people with close characteristics.

How often an individual or some individuals travel and exhibit some specific patterns?

In order to figure out how often an *Individual* or some *Individuals* perform some travels and then exhibit *TravelPatterns* are key issues to derive displacement regularities in the city. Pattern or irregular *Trajectories*, as derived from frequent and similar patterns, can reflect travel habits in the city while irregular patterns might reflect some outlier processes. Such inputs should be of great value to provide a better picture of travel patterns in the city, this being of interest for transportation planning. In order to implement such analysis, aggregations based on semantic, temporal and spatial criteria should be applied. Travel regularities and irregularities should be derived and classified according to different time periods (for instance, the travel displacement of individuals during week days or weekends are usually worth being distinguished).

Why individuals perform some travels patterns?

Travel patterns, either regular or irregular, can be generated and influenced by several factors. Traffic jams are common examples of travel patterns, either regular or irregular, this being a direct consequence of the aggregation of a convergence in space and time of travel choices made by individuals. For example, people tend to go to city centres for shopping in the weekend, this generating typical travel patterns and huge crowds in the city centre. Exploring why people exhibit certain travel patterns can provide a valuable input to traffic management and transportation planning in the city. Such travel patterns should be explored. Universal queries tend to discover universal phenomenon that happens in specific times and places.

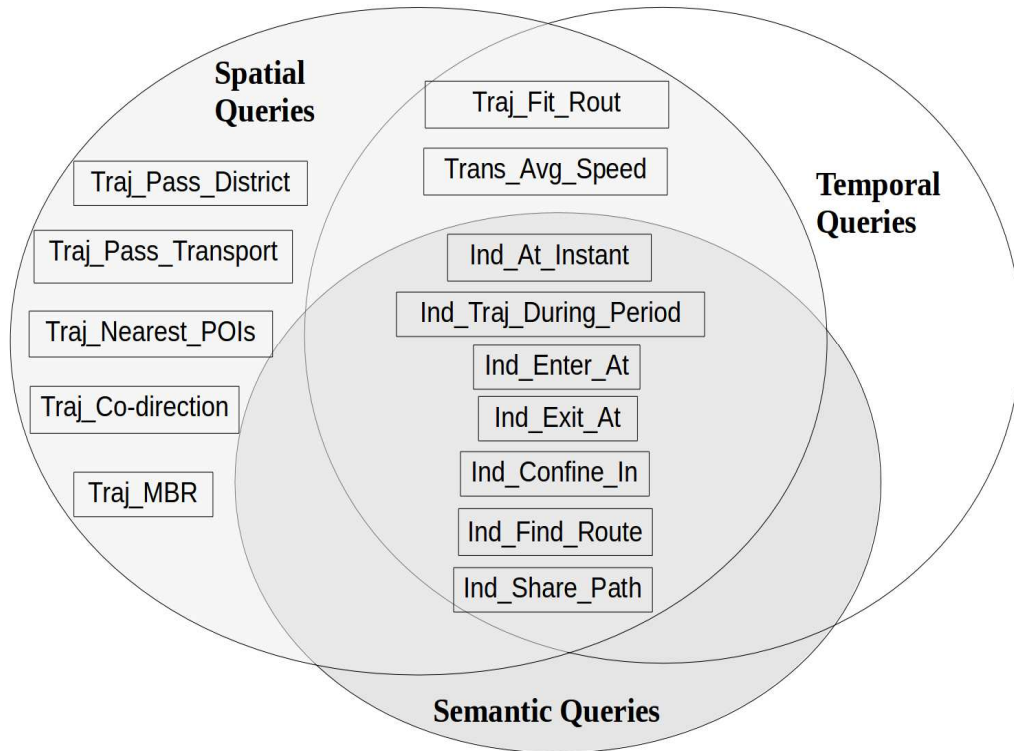


Figure IV. 1. Query Categories

What are the preferred travel modes used by individuals when performing some travel patterns?

TravelPatterns include different travel modes (e.g., walking, taking a bus, driving...), travel routes and time durations. Therefore, different criteria can be applied in order to evaluate the degrees of similarities of *TravelPatterns* in space and time. In particular, travel modes should provide some useful insights to the comprehension of travel habits, *TravelPatterns* and *Trajectory* similarities that emerge in the city.

What are(is) the most frequent route(s) taken from one destination to another one?

For large cities such as the ones considered by our study, many *Routes* are likely to appear, in particular when relating one place to another one. Not only frequent visited places can be identified, but also the most common *Routes* taken

when moving from one place to another one. By searching for such common *Routes*, and comparing them, potential travel patterns or even traffic problems might be identified.

IV.3.2 Data Analysis Functions and Queries

Taking the *Travel Displacement Model* and the category of travel displacement queries as the basic foundation of our approach, functions and queries are designed for discovering travel patterns in the city. The design of the functions and queries are based on the three categories discussed above. Considering the potential complexity of any travel displacement analysis, most of the functions and queries identified combine the spatial, temporal and semantic dimensions. Note that most of the functions and query examples presented are related to some data explorations rather than data manipulation tasks such as insertion, updating and delete. Let us introduce the different categories of functions and queries to design.

(1) Functions

Functions constitute basic components of search queries. A spatial database should usually offer a series of operators and functions that mainly deal with location-based criteria, either quantitative (e.g., metrics) or qualitative (e.g., topological). Nowadays, many basic spatial operators and functions are most often available in current spatial database system, such as Oracle Spatial, PostGIS etc. These functions include metrics and geometry constructors (e.g., distance, length, buffer), topological operators (e.g., cross, within, touch). However, such operators and functions are not always sufficient for specific applications with high-level semantics such as for the analysis to be conducted by our study oriented to the analysis of travel displacements. At least, some specific functions should be derived and aggregated with some explicit semantics that will fit the functions required by our approach. Such functions should directly and explicitly search form some people's travel displacement, and should be built to deal with spatial, temporal and semantic data.

The design of the functions is partly adapted from the taxonomy of spatio-temporal operators introduced in a related work (Wu & Claramunt, 2014). The main spatio-temporal topological concepts considered by the function design are given in Table IV.12.

Table IV. 12. Spatio-temporal Operators, adapted from Wu & Claramunt (2014)

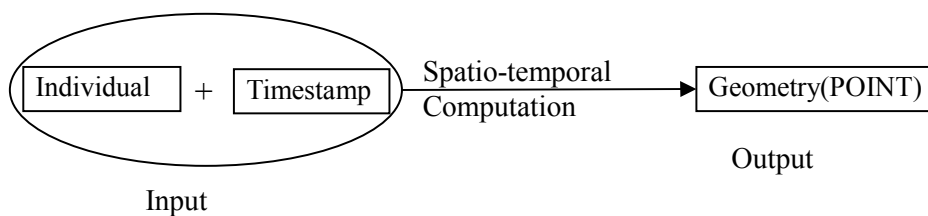
ST Operator	Signature	Output	Description
Enter(A,B, T)	Trajectory Geometry x	Boolean	Enter(A,B,T) denotes the case where the trajectory A starts on the exterior of B then move to the interior of B during a time interval T.
Exit(A,B,T)	Trajectory Geometry x	Boolean	Exit(A,B,T) denotes the case where the trajectory A starts on the interior of B and then move to the exterior of B during a time interval T.
PassThrough(A,B,T)	Trajectory Geometry x	Boolean	PassThrough(A,B,T) denotes the case where the trajectory A starts on the exterior of B then move to the interior of B and then move to the exterior of B.

When applied to the human travel displacement data model, the Geometry can be either a Polyline or a Region.

Function 1 Ind At Instant

The function Ind_At_Instant is a spatio-temporal function whose objective is to find out where is a given *Individual* at a given time instant. The function takes as an input an *Individual* and a time instant and returns a geometry point that denotes the location of that *Individual* at that time.

Figure IV. 2. Function1 Ind_At_Instant



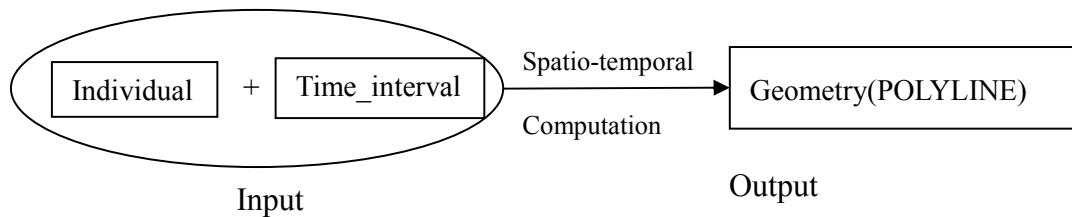
Ind_At_Instant: Individual x Timestamp-> Geometry(POINT)

Example Query 1:

“Where were Wang Hong at 8:00 am on the 13th February, 2009?”

Select Ind_At_Instant(id,'2009-02-13 08:00:00') from IndividualTable where name = 'Wang Hong';

Figure IV. 3. Function 2 Ind_Traj_During_Period



Function 2 Ind Traj During Period

The function Ind_Traj_During_Period searches for the part of the *Trajectory* performed by a given *Individual* within a given time interval. The input of the function is an *Individual* and a time interval. The output of the function is a geometry *Polyline* representing the part of the *Trajectory* within that given time interval.

Ind_Traj_During_Period: Individual x Time_Interval -> geometry(POLYLINE)

Example Query 2:

“Did Li Ming and Li Lei share some common trajectory parts while travelling from 5pm to 7pm on the 15th July, 2010?”

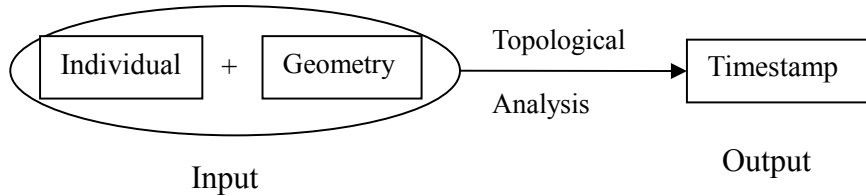
Select intersects(Ind_Traj_during_period(indiv1.id, '2010-07-15 17:00:00', '2010-07-15 19:00:00'), during _period(indiv2.id, '2010-07-15 17:00:00','2010-07-15 19:00:00')) from IndividualTable indiv1, IndividualTable indiv2 where indiv1.name = 'Li Ming' and indiv2.name = 'Li Lei';

Function 3 Ind Enter At

Traj_Enter_At is a spatio-temporal function that searches for the time instant when possibly a given *Individual* enters a spatial entity denoted by a geometry. This function has as an input the individual and a geometry, and returns a timestamp which is the time at which the given *Individual* enters the given geometry.

Ind_Enter_At: Individual x Geometry-> Timestamp

Figure IV. 4 Function 3 Ind_Enter_At



Example Query 3:

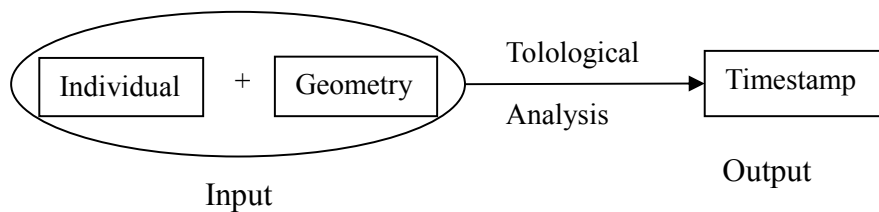
Find out when did the individual called “Li Ming” entered the road called “Tianjin Road”?”

```
Select Ind_Enter_At(indiv.id, trans.geom) from IndividualTable indiv,
TransportationSegmentTable trans where trans.type = 'road' and trans.name =
'Tianjin Road' and indiv.name = 'Li Ming';
```

Function 4 Ind_Exit_At

The function Ind_Exit_At is a sort of a reverse one of the function Ind_Enter_At. The objective is to search for the time instant when the given *Individual* possibly leaves a given geometry. The input and output of the function are similar to the function Ind_Enter_At.

Figure IV. 5. Function 4 Ind_Exit_At



Traj_Exit_At: Individual x Geometry-> Timestamp

Example Query 4:

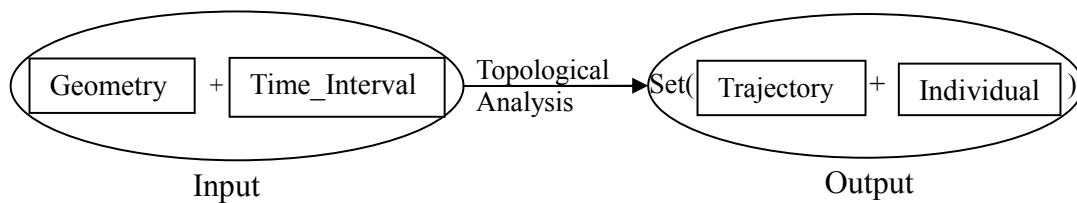
“At what time Li Lei entered in the District Haidian, and at what time he left that District on 15 July, 2010?”

```
Select Ind_Exit_At(indiv.id, district.geom)-Ind_Enter_At(indiv.id, district.geom)
from IndividualTable indiv, DistrictTable district where indiv.name = 'Li Lei' and
district.name = 'Haidian' and Ind_Enter_At(indiv.id, district.geom) > '2010-07-15
00:00:00' and Ind_Exit_At(indiv.id, district.geom) < '2010-07-15 23:59:59';
```

Function 5 Ind_Confine_In

The function *Ind_Confine_In* searches for the individual *Trajectories* located in a spatio-temporal Minimum Bounding Rectangle (MBR) for a given time interval. The input of the function is a Geometry polygon with a time constraint. The output of the function gives a list of pairs of *Individual* and *Trajectories*.

Figure IV. 6. Function 5 *Ind_Confine_In*



Ind_Confine_In: Geometry(Polygon) x Time_Interval -> Set(Trajectory, Individual)

Example Query 5:

“Find all the *Trajectories* and individuals that passed through the region materialised by the Minimum Bounding Rectangle BOX2D('116.3027', '39.9310', '116.3030', '39.9315') between 7am and 9am on 14th July, 2010 ;”

```
Select Ind_confine_in(ST_MakeBox2D(ST_POINT(116.3027, 39.9310),
ST_POINT(116.3030, 39.9315) :: geometry), '2010-07-14 07:00:00', '2010-07-14
09:00:00');
```

Function 6 Traj_MBR

The function *Traj_MBR* is a spatial function that returns the MBR that a given *Individual(s)* passed through while performing one or more trajectories. With input

as one-to-many *Trajectories*, the function returns a MBR that contains all of the input *Trajectories*.

Traj_MBR: set(Trajectory) -> Region

Example Query 6:

“Find the range of trajectories "Li Ming" performed in one given week.”

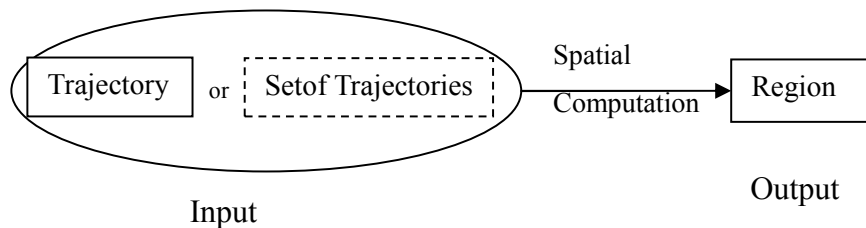
```
Select Traj_MBR(select array(select trajid from TrajectoryTable traj,
IndividualTable indiv where traj.individualid = indiv.id and indiv.name = "Li Ming"
and starttime between '2009-04-05' and '2009-04-12'));
```

Example Query 7:

“Find the range of trajectories that passed close to the POI building "Sohu" between 17:30 and 18:30 on 17th November, 2010.”

```
Select Traj_MBR(select array(select traj.id from TrajectoryTable traj,POITable poi
where poi.name = "Sohu" and poi.id in (select traj.id from NearestPOI(startpt)) and
starttime between '2010-11-17 17:30:00' and '2010-11-17 18:30:00'));
```

Figure IV. 7. Function 6 Traj_MBR



Function 7 Traj Pass District

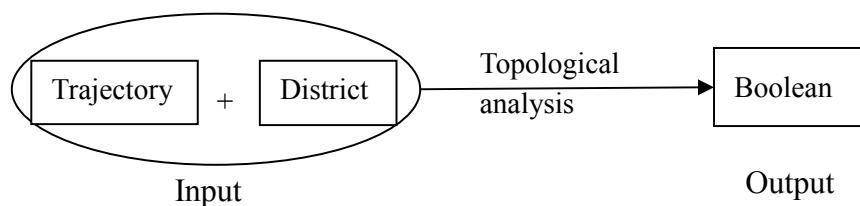
The function *Traj_Pass_District* finds out if a given *Trajectory* has passed by a given *District*. The action *Pass* applies the *PassThrough* function introduced in Table IV.12 and applied to a given *Trajectory* and a given *District*. The function takes as an input a *Trajectory* and a *District* and returns a Boolean value, that is, "TRUE" if the *Trajectory* passed by the *District*, "FALSE" on the contrary.

Traj_Pass_District: Trajectory x Polygon -> Boolean

Example Query 8:

“Find the trajectories that cross the District Chaoyang identified with an Id 335”

Figure IV. 8. Function 7 Traj_Pass_District

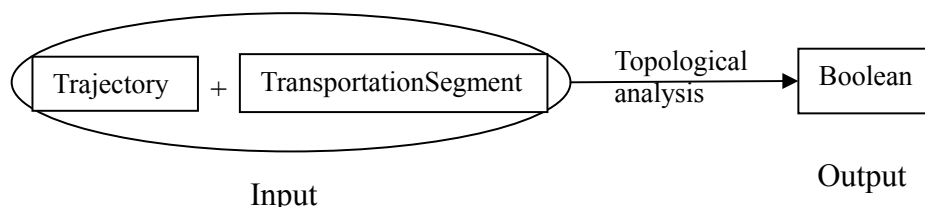


Select trajectory.id, starttime, endtime from TrajectoryTable traj, District district where district.id = 335 and Traj_Pass_District(traj.id, district.geom);

Function 8 Traj_Pass_Transport

The function *Traj_Pass_Transport* determines if a given *Trajectory* has passed through a named *TransportationSegment*. As for the function *Traj_Pass_District*, the function *Traj_Pass_Transport* also uses the topological relationship *PassThrough*. The function takes as an input a *Trajectory* and a *TransportationSegment* and returns a Boolean, "TRUE" if the *Trajectory* passed by the *TransportationSegment(s)* with the given name of the transportation line, "FALSE" on the contrary.

Figure IV. 9. Function 8 Traj_Pass_Transport



Pass_Transport: Trajectory x (Polyline) -> Boolean

Example Query 9:

“Find all the trajectories that passed through the road “西二环” on 6th Sept. 2009.”

Select trajectory.id, starttime, endtime from TrajectoryTable traj, TransportationSegmentTable trans where trans.name = ‘西二环’ and tran.type = ‘road’ and Traj_pass_transport(traj.id, trans.geom) and date(traj.starttime) = ‘2009-09-06’;

Example Query 10:

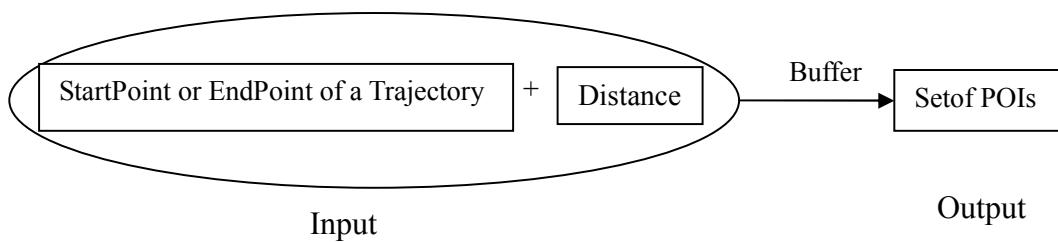
“When did the Individual named “Shu Ji” traveled on metro line No. 2?”

```
Select extract(date from starttime) from TrajectoryTable traj,
TransportationSegmentTable trans, IndividualTable indiv where traj.individualid =
indiv.id and Traj_pass_transport(traj.id, trans.geom) and trans.name = ‘No. 2’ and
trans.type = ‘metro’;
```

Function 9 Traj_Nearest_POIs

The Traj_Nearest_POIs is a function that searches for the points of interest close to a given point. This can be for instance either the startpoint or endpoint of a given Trajectory. Two additional and embedded functions *StartTraj* and *EndTraj* are defined in order to give the start point and end point of the given trajectory, respectively.

Figure IV. 10. Function 9 Traj_Nearest_POIs



Functions signature:

Traj_Nearest_POIs: Point x Distance -> set(POI)

StartTraj: Trajectory -> Point

EndTraj: Trajectory -> Point

The function Nearest_POIs searches for the POIs within a given distance from a point and that can be derived from the start or end points of a given Trajectory as follows.

Traj_Nearest_POIs: StartTraj/EndTraj(Trajectory) x Distance -> set(POI)

The function result gives a set of POIs.

Example Query 11:

“Find the points of interests within a distance of 50 meters from the starting point of a trajectory ‘Li Ming’ performed in the morning of 31st Oct. 2008.”

```
Select * from Traj_Nearest_POIs ((select StartTraj from TrajectoryTable traj, IndividualTable indiv where indiv.name = 'Li Ming' and traj.individualid = indiv.id and starttime between '2008-10-31 06:00:00' and '2008-10-31 11:00:00'), 50);
```

Example Query 12:

“Search for POIs that are within a 50 meters’ distance to the destinations of the trajectories performed by “Zhang Hua”, and also the ones which are within a 50 meters’ distance to the start of the travel of “Huang Shan” in the afternoon of 12 May, 2010”

```
select * from Traj_Nearest_POIs ((select endpt from TrajectoryTable traj, IndividualTable indiv where indiv.id=traj.individualid and indiv.name = “Zhang Hua” and endtime between '2010-05-12 14:00:00' and '2010-05-12 18:00:00'),50) poi_1 JOIN Traj_Nearest_POIs((select startpt from TrajectoryTable traj, IndividualTable indiv where indiv.id=traj.individualid and indiv.name = “Huang Shan” and endtime between '2010-05-12 14:00:00' and '2010-05-12 18:00:00'),50) poi_2 on poi_1.id = poi_2.id;
```

Example Query 13:

“Find the POI types' distribution for the ones which are within a 50 meters distance to the starting point of the trajectory id 23.”

```
Select count(id),type from Traj_Nearest_POIs((select startpt from trajectory where id = 23),50) group by type;
```

Function 10 Traj Direction

The function *Traj_Direction* derives the direction of a given *Trajectory*. We define the direction of a *Trajectory* as the true bearing in degree (North-based azimuth as the angle in radians measured clockwise) from the starting point of the *Trajectory* to the end point of the *Trajectory*. That is to say that we consider the direction from the starting point to the end point as the direction of travel. This information can summarize the overall direction of an individual *Trajectory*, and thus allowing us to study the moving and directional flow in the city at a macroscopic and aggregated level.

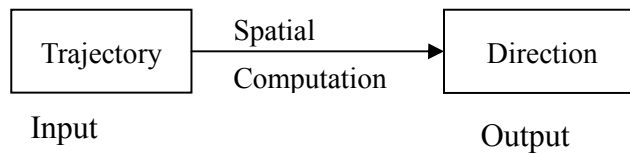
Traj_Direction: Trajectory -> Bearing

Example Query 14:

“What is the travel direction of Wang Lu’s trajectory around 18:00 of 16th July, 2009?”

Select Traj_Direction(traj.id) from TrajectoryTable traj, IndividualTable indiv where indiv.id = traj.individualid and indiv.name = "Wang Lu" and starttime between '2009-07-16 17:30:00' and '2009-07-16 18:30:00';

Figure IV. 11. Function 10 Traj_Direction



Example Query 15:

“Find the individual(s) ids and names whose travel direction is approximately from the North to the South (10 degree of error allowed) between 6:00am and 10:00am on 6th June, 2011.”

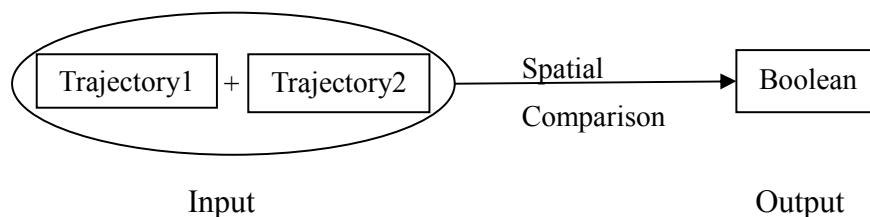
Select indiv.id, indiv.name from IndividualTable indiv, TrajectoryTable traj where indiv.id = traj.individualid and Traj_Direction(traj.id) between 170 and 190 and traj.starttime between ‘2011-06-06 06:00:00’ and ‘2011-06-06 10:00:00’;

Function 11 Traj Co-direction

The Traj_Co-direction function evaluates if two given *Trajectories* are moving along a similar direction. This function is implemented with an approximation parameter given in degrees, that is, when the heading of two given *Trajectories* is lower than 10 degrees we consider that the two *Trajectories* considered have a similar direction.

Traj_Co-Direction: Trajectory x Trajectory -> Boolean

Figure IV. 12. Function 11 Traj_Co-direction



Example Query 16:

“Return the trajectories that have similar directions between 7:00am and 9:00am on the 9th April, 2009.”

```
Select traj1.id, traj2.id from TrajectoryTable traj1, TrajectoryTable traj2 where
Traj_Co-direction(traj1.id, traj2.id) and (traj1.starttime between '2009-04-09
07:00:00' and '2009-04-09 09:00:00') and (traj2. starttime between '2009-04-09
07:00:00' and '2009-04-09 09:00:00') and traj1.id <> traj2.id;
```

Example Query 17:

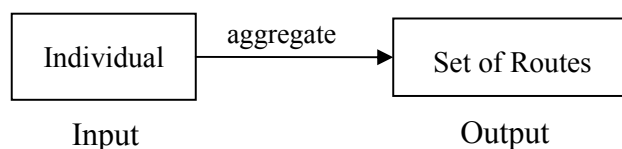
“Did the individual named ‘Li Ming’ travel along the same direction from 7:00am to 9:00am on the 10th April, 2009 and 9th April, 2009?”

```
select Traj_Co-Direction(traj1.id, traj2.id) from TrajectoryTable traj1,
TrajectoryTable traj2, IndividualTable indiv where (traj1.starttime between '2009-
04-09 07:00:00' and '2009-04-09 09:00:00') and (traj1.starttime between '2009-04-10
07:00:00' and '2009-04-10 09:00:00') and traj1.individualid = indiv.id and
traj2.individualid = indiv.id and indiv.name = ‘Li Ming’;
```

Function 12 Ind Find Route

The *Ind_Find_Route* is a function whose objective is to extract the *Route(s)* performed by one-to-many *Individuals*. According to the Travel Displacement schema, *Route* is an entity that contains a list of *TransportationSegments* with their respective *Starttimes* and *Endtimes*. A *Route* represents the travel path of one-to-many *Individuals* that have repeated this *Route* at similar hours but possibly in different days. The function *Ind_Find_Route* has as an input one-to-many *Individuals* and as an output a set of *Routes*.

Figure IV. 13. Function 12 *Ind_Find_Route*



Ind_Find_Route: Set(Individual) -> Set(Route)

Example Query 18:

“Find all the Routes performed by “Xiaoming.”

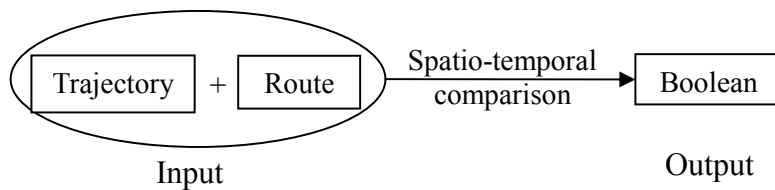
```
Select * from Traj_Find_Route((select id from IndividualTable where name = 'Xiaoming'))
```

Function 13 Traj Fit Route

The function *Traj_Fit_Route* determines whether a given *Trajectory* pass through a given *Route*. As mentioned above, a *Route* is defined as a set of *TransportationSegments* that a given *Individual* repeats several times. The *Routes* considered represent the *TravelPatterns* collected from the history data of a given *Individual*.

Traj_Fit_Route: Trajectory x Route -> Boolean

Figure IV. 14. Function 13 Traj_Fit_Route



Example Query 19:

“What are the regular travel paths of the trajectories that started today?”

```
Select route.path,route. from TrajectoryTable traj, RouteTable route where Traj_Fit_Route(traj.id, route.id) and extract(date from traj.starttime ) = currentdate();
```

Example Query 20:

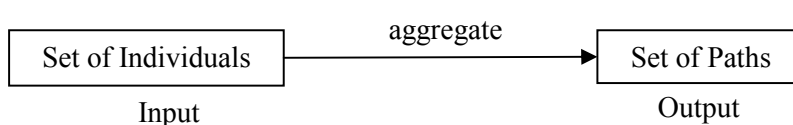
“Did the individual named Liu trajectory follow the route id 2 today?”

```
Select Traj_Fit_Route(traj.id) from TrajectoryTable traj, RouteTable route, individual indiv where route.id = 2 and date(traj.starttime) = currentdate() and indiv.id = traj.individualid and indiv.name = 'Liu';
```

Function 14 Ind Share Path

The function *Ind_Share_Path* searches for the *Path(s)* that the *Individual* has already repeated more than once before. A *Path* is defined as a list of *TransportationSegments* that the *Individual* has repeated while performing several *Trajectories*. A *Path* is an attribute of the object type *Route*. This function returns the *Path(s)* that one-to-many *Individual* passed by several times.

Figure IV. 15: Function 14 Ind_Share_path



Ind_Share_Path: set(individual) -> set(path)

Example Query 21:

“Find similar paths previously performed by Li Lei and Wang Hong.”

Select *Ind_Share_Path*(array(a.id,b.id)) from *IndividualTable* a, *IndividualTable* b where a.name = 'Li Lei' and b.name = 'Wang Hong'

Example Query 22:

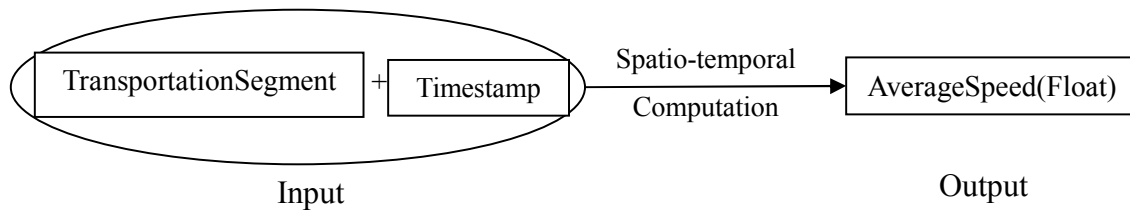
“Find the paths that Li Lei has repeated during the week from 7th August, 2008 to 14th August, 2008 ”

Select *Ind_Share_Path*(indiv.id) from *TrajectoryTable* traj,*IndividualTable* indiv where traj.individualid = indiv.id and indiv.name = 'Li Lei' and traj.starttime > '2008-08-07' and traj.endtime <'2008-08-14')) ;

Function 15 Trans Avg Speed

The function *Trans_Avg_Speed* derives the average speed of a *TransportationSegment*. The average speed of a *TransportationSegment* reflects the traffic conditions for a given temporal interval. It can reflect individuals’ travel decisions as well as it can be considered as a reference for transportation planning tasks.

Figure IV. 16. Function 15 Trans_Avg_speed



Trans_Avg_Speed: set(TransportationSegments) x Timestamp -> Speed

The function *Trans_Avg_Speed* has one-to-many *TransportationSegments* as input and a speed as an output.

Example Query 23:

“Find the average speed of the segments of Beijing road at 12 am.”

Select Trans_Avg_Speed(id, '12:00:00') from TransportationSegment where name = 'Beijing Road' and type = 'road';

Example Query 24:

“Find the TransportationSegment with the minimal speed at 6pm.”

Select * from TransportationSegment where Trans_Avg_Speed(id, '18:00:00') in (select min(Trans_Avg_Speed(id, '18:00:00')) from TransportationSegment);

IV.4 DISCUSSION

This chapter introduces a logical approach to provide query capabilities that will favour further analysis of travel displacements. The whole approach has been implemented on top of the conceptual representation introduced in the previous chapters. In order to explore and validate the manipulation capabilities and opportunities given by the travel displacement model, a series of data manipulation and queries have been identified and classified.

The database developed at the logical level covers most if not all the modelling abstractions identified by the travel displacement model, including object types and their attributes, as well as the relationships that related them. Every modelling object identified by the conceptual model is likely to be mapped towards a relational table

in the implemented database. This database contains spatial, temporal, spatio-temporal and semantic data types. A thoroughly implementation of the human travel displacement model together with the development and implementation data manipulation functions and mechanisms provide a relevant support to analyse and mine the database population implemented. In order to not only model some of the most common queries to apply, but also to keep track of the different query possibilities that present a regular interest at the implementation level, a series of data manipulation functions have been designed and implemented as primitive query operators to be applied on the data types implemented at the logical level. The design of the different functions are organised through the notion of semantic, spatial and spatio-temporal categories. Overall, such functions deal with either basic or complex spatio-temporal queries. Indeed, combination of such functions including for example basic derivations and calculations, specific spatial data manipulation functions dealing with topological, cardinal and metric operations. A series of query examples illustrate the whole potential of the approach, and how either explicit functions or implicit data mining operations can help to understand some human displacement patterns in the city. The case study developed and applied to the city of Beijing illustrates the different categories of functions implemented on top of real and large urban trajectories can be. Last but not least, the whole set of queries illustrated in this chapter show how useful such a modelling approach can be for the analysis of human displacements in a given city.



Implementation & Experimental Evaluation

V.1	IMPLEMENTATION OBJECTIVE	86
V.2	EXPERIMENTAL DATA SET	86
V.3	COMPUTING AND SOFTWARE ENVIRONMENTS.....	89
V.3.1	DEVELOPMENT ENVIRONMENT.....	89
V.3.2	DATABASE ENVIRONMENT	89
V.4	FRAMEWORK CONSTRUCTION	90
V.4.1	DATA PRE-PROCESSING.....	91
V.4.1.1	<i>Construction of the Transportation Network.....</i>	<i>91</i>
V.4.1.2	<i>Map Matching</i>	<i>99</i>
V.4.2	DATABASE CONSTRUCTION AND ANALYSIS	102
V.4.2.1	<i>Data Structure Generation.....</i>	<i>102</i>
V.4.2.2	<i>Route Extraction Algorithm.....</i>	<i>105</i>
V.5	RESULTS AND DISCUSSION	107

V.1 IMPLEMENTATION OBJECTIVE

This chapter introduces the experimental implementation and developments made to validate the potential of the travel displacement model identified at the conceptual level and illustrated in the previous chapter. At the implementation level, and in order to study the feasibility and performance of our modelling approach, a large and real urban trajectory dataset derived from GPS data by the Geolife project developed in the city of Beijing has been used as a testbed (Zheng et al. 2008, 2009, 2010). Most of the implementation tasks have been performed on top of the PostgreSQL/PostGIS database system as this choice offers robust functions to integrate and manipulate various data types as well as sufficient extensibility to define and specify additional manipulation functions at the query level. The objectives of this experimental experimentation are as follows:

- To evaluate the feasibility of the mapping of the conceptual transportation displacement model to the logical level,
- To implement the spatial, temporal and thematic functions at the query language level,
- To experiment the potential of the transportation displacement model when applied to a large trajectory dataset,
- To implement the modelling approach on top of an existing spatial database,
- To explore to which degree the application of the framework to a large trajectory dataset can reveal some displacement patterns in the city.

V.2 EXPERIMENTAL DATA SET

The experiments developed have been realized on top of a very large GPS trajectory dataset collected by the Geolife project and sponsored by Microsoft Research Asia (Zheng et al. 2008, 2009, 2010). This dataset represents the urban

trajectories performed by 178 users over a period of four years from April 2007 to October 2011. A GPS trajectory is represented by a sequence of time-stamped points, and whose geographical attributes are given in the three spatial dimensions, that is, latitude, longitude and altitude. Overall, this dataset contains 17,621 trajectories with a total distance covered of 1,251,654 kilometres and a total duration of 48,203 hours. These trajectories were recorded by different GPS loggers and GPS-phones, and have a variety of sampling rates. 91 percent of the trajectories are logged in a dense representation, e.g. every 1~5 seconds or every 5~10 meters per point. Indeed, this dataset available is mainly made of spatio-temporal data where trajectory locations and user *ids* constitute the main information available. However, as it will be shown in the following sections, the idea behind the experiments and queries developed are to extract some displacement patterns from such a large dataset.



Figure V. 1. Geolife Dataset

This dataset encompasses a large set of inhabitants' trajectories and whose transportation modes are part of the semantics associated. While some of the users have carried a GPS logger for a few years, others only have used a similar setup for few weeks. The logical data structure of the dataset is as follows.

Table V. 1. Logical representation of the trajectory dataset

Field Name	Data Type	Length	Description
ID	int	4	Trajectory id
Longitude	float	8	GPS-based Longitude
Latitude	float	8	GPS-based Latitude
Altitude	float	8	GPS-based Altitude
DatebyNum	float	8	Number of days passed since Dec 30 1899
Date	varchar	10	Date
Time	varchar	8	Time

Table V. 2. Transportation mode logical representation

Field Name	Data Type	Length	Description
TrajectoryID	int	4	Trajectory id
Start_time	Datetime	19	Timestamp (with date and time)
End_time	Datetime	19	Timestamp (with date and time)
Transportation Mode	varchar	8	Transportation mode of the trajectory segment

Additional geographical data from Beijing city are available and integrate roads, railways, buildings, points of interest, land-use, buildings, waterways, natural features etc. These basic geographic datasets of the city of Beijing are also essential

to our experiment, as travel displacements are not completely independent from the city environment. The transportation data and other spatial data of the city of Beijing City came from public domain spatial datasets, including the road and subway networks, districts, and a series of POIs extracted from the Baidu Map service (Lbsyun Baidu, 2016).

V.3 COMPUTING AND SOFTWARE ENVIRONMENTS

V.3.1 Development Environment

The implementation of the framework is performed on a computer with a processor Intel Core i7 and a memory size is 8 GB. The operating system is CentOS 7.2.1.

Table V. 3. Computer Environment

Processor	Inter Core i7
Memory	8 GB
Hard Drive	1 TB
Operating System	CentOS 7.2.1

V.3.2 Database Environment

The implementation of the urban displacement model and the instantiation of the database and functions are built on top of PostgreSQL/PostGIS. PostgreSQL is an open source object-relational database system based on POSTGRES, Version 4.2, developed at the Computer Science Department of the University of California at Berkeley (Douglas 2003). It has full support for foreign keys, joins, views, triggers, and stored procedures. It includes most SQL2008 data types, usual alphanumerical and temporal data types. It also supports storage of binary large objects, including pictures, sounds and videos. The extensibility of PostgreSQL allows the users to customize functions and operations, which meet the goal of implementing the designed functions developed in the previous chapter.

PostGIS is an extension to the PostgreSQL object-relational database system which allows geographical data types to be specified in the database (Ramsey 2004, 2016). PostGIS includes support for GiST-based R-Tree spatial indexes, and functions for the analysis and processing of GIS objects.

PostGIS supports geometry data types for Points, LineStrings, Polygons, MultiPoints, MultiLineStrings, Multipolygons and GeometryCollections, as well as a raster data type. PostGIS also supports wide range of spatial operations and functions, including

- 1) spatial predicates for determining the interactions of geometries using the 9-intersection model proposed by Egenhofer (1990) (provided by the GEOS software library, which is a C++ library that supports basic geometry manipulations);
- 2) spatial operators for deriving basic metrics such as area, distance, length and perimeter;
- 3) specific operators for geospatial set operations such union, difference, symmetric difference and buffers (provided by GEOS).

In fact PostGIS has the advantage of facilitating the implementation of our conceptual model as well as the development of built-in functions. The geometry data type offered by PostGIS is embedded in the attributes of the spatial or spatio-temporal objects. Several spatial and topological operators have been embedded in the functions defined, as well as basic operators of the queries illustrated in the next sections. For example, `ST_Buffer`, which is a function that returns a region that covers all points within a given distance from an input geometry, is applied to search for the POIs around a given trajectory in combination with another function the `ST_Intersects`, which determines if two geometries intersect each other.

V.4 FRAMEWORK CONSTRUCTION

In order to evaluate the practicability of the model and functions presented in Chapter III and Chapter IV, an integrated construction approach is necessary. The construction of human displacement framework is mainly composed of several algorithms and database construction SQL and Pl/PgSQL command designed.

The following introduces the essential algorithms to construct the framework.

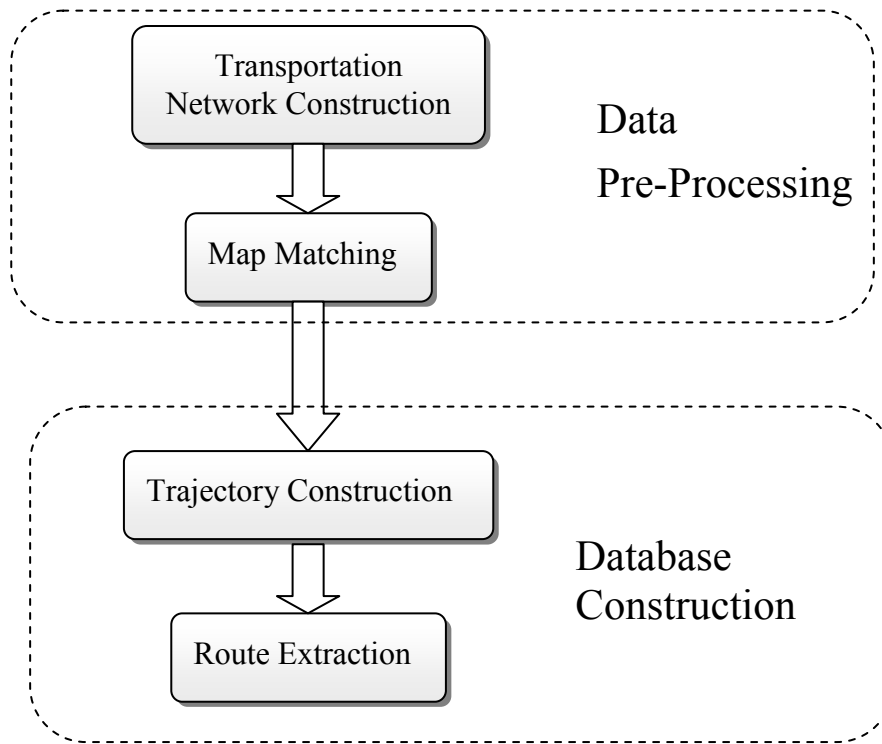


Figure V. 2 Processing and Construction Procedure

V.4.1 Data Pre-Processing

The real trajectory data used as a testbed is obtained from GPS devices or GPS equipped citizens' smartphones while the underlying transportation network is made of roads and railway lines of Beijing City. Such real data is usually in the form of raw data that requires some pre-processing in order to be formatted according to some predefined regarding all spatial, temporal and semantic dimensions. This is particularly important and the case for a prior matching to our conceptual model. Not only the trajectory data should be mapped to the displacement model but the underlying geographical data of Beijing city should also follow a similar process. The following subsections introduce the pre-processing steps and algorithms applied to the incoming trajectory data in order to associate them to the transportation network.

V.4.1.1 Construction of the Transportation Network

Human beings in the city do not move freely in a two-dimensional space but are rather constrained by spatially embedded networks (e.g., streets, paths) (Güting, 2006). Such network is considered as one of the spatial constraints identified by our

modelling approach. The usual model largely used for representing an urban network is based on topological graph theory (Gross and Tucker, 1987). Accordingly, a network is modelled as a directed graph $G=(V,E)$, where V denotes the set of nodes and $E \subseteq V \times V$ the set of edges. In most transportation network applications, a vertex represents a road intersection or an end road, while an edge is a connection between two vertices (Claramunt & Mainguenaud, 1996; Chaudhry & Mackaness, 2006). Additional modelling abstractions are often derived from the basic structure of a transportation network. For instance, paths and routes have been widely used and modelled in related work (Claramunt & Mainguenaud 1999). In another related work, Etches et al. (1999) developed an object-oriented GIS model of a traffic system. The basic elements of an urban network are modelled as objects characterised in the spatio-temporal and semantic dimensions, the objective being to relate such abstraction to traffic flows and to analyse emerging patterns at different levels of abstraction. Bi-directional road segments and nodes are topologically connected, and real-time associated to traffic data. A series of network constraints such as every road is terminated by two nodes are also predefined.

Inspired by principles identified by the object-based traffic suggested by Etches et al. (1999), the transportation network module is based on two elementary modelling abstractions: *TransportationSegment* and *TransportationNode* are topologically connected. *TransportationSegment* is defined as a non-oriented spatial polyline while *TransportationNode* is defined as a punctual primitive. The spatial and non-spatial attributes of *TransportationSegment* includes *Length* (i.e., length of the segment), *Name*, and every *TransportationSegment* is linked to two nodes denoted as *Startnode* and *Endnode*.

The original transportation network of the city of Beijing has been vectorised from a scanned map, this generating a very large dataset of roads and railways (railways are omitted in our experiment as railways stops are not materialised).

A connectivity graph should be then constructed. The idea is to extract vertices from streets, and then to generalise a network graph. When considering our modelling approach, all successive *TransportationSegments* must be connected (i.e., the ones that successively represent a route), that is, for all *TransportationSegments* either the starting *TransportationNode* or the ending *TransportationNode* of that *TransportationSegment* is either the starting or ending *TransportationNode* of at least another *TransportationSegment*. The construction of the transportation network of Beijing city is made by a topological graph-based algorithm. Although a Graph

class was implemented and available in Java, and where *Edge* and *Node* and the two object types available in Oracle, there is still a need for some user-designed methods to create and build up the graph or network.

After the basic topological checking, there should be no topological mistakes in the incoming transportation data (i.e., no dangles referring to a polyline having an end node that is not connected to another polyline, no overlap, no pseudo nodes, no self-overlap). A connectivity graph is derived and stored in form of matrix suggested by Jiang & Claramunt (2004), where a node represents a road intersection, and a road segment denotes the part of the road network between two nodes.

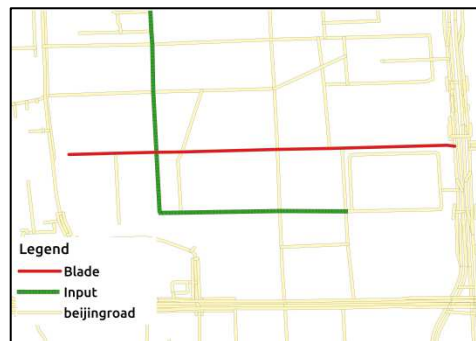


Figure V. 4. Blade Crosses Input

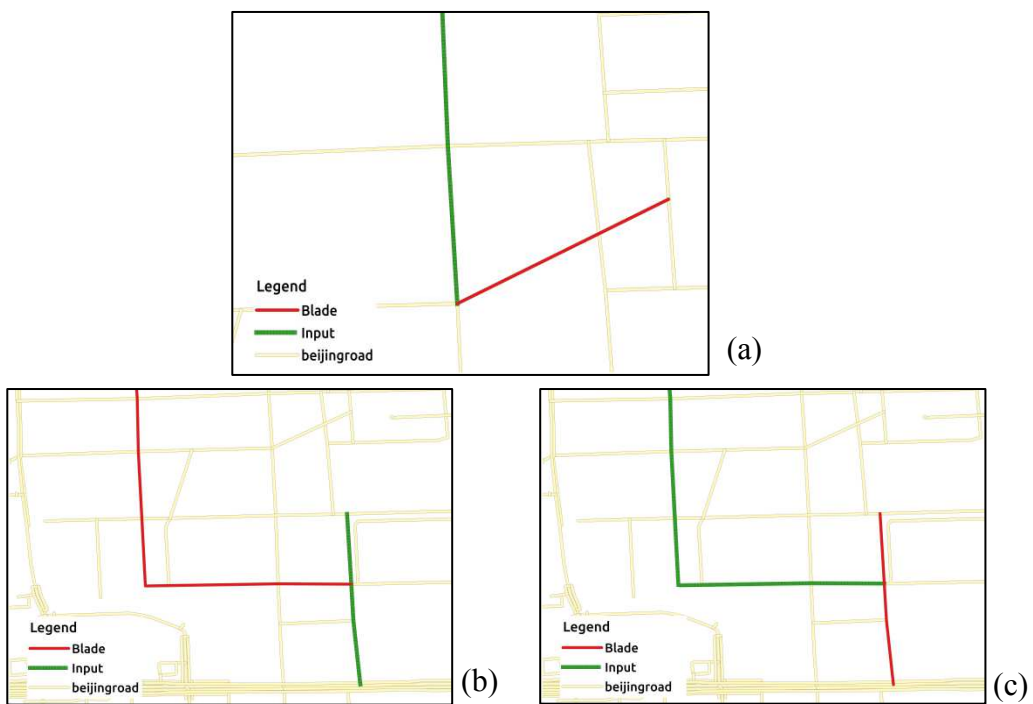


Figure V. 3. Blade touches Input

Algorithm 1 create-connectivity-graph // V is a set of streets; assume that street ID range from 0 to $n-1$; R is the output matrix of the connectivity graph

```

Begin
   $R \leftarrow [0 \dots n-1][0 \dots n-1]$  // an empty matrix with all elements as null values
   $V' \leftarrow V$  // target set of vertices
  for every  $v \in V$  do
    for every  $w \in V$  do
      if INTERSECTION ( $v, w$ ) then
         $R[v][w] \leftarrow 1$ 
      else
         $R[v][w] \leftarrow 0$ 
      end if
    end for
  end for
end create-connectivity-graph

```

The Algorithm 1 offers a proper method to extract materialised intersections from the urban network. However, some real intersections might be not apparent in the underlying urban network and should be identified in order to provide an accurate geometrical reference. A second algorithm is then designed and whose objective is to identify such remaining road intersections. The principle of this second algorithm is as follows. Every road segment is considered as an input and every road segment that intersects this input (Figure V.3) and not currently identified as such by Algorithm 1 is considered as a candidate blade road segment. Next, the vertices that denote these intersections are extracted and split accordingly the road segments. This second algorithm has to also take into account the complexity of the underlying city network. For example, there might be more than one intersection point between a given input road segment and a candidate blade road segment (Figure V.5(b)). These different intersection cases should be identified. The intersection cases to identify between an input road segment and a candidate blade can be generalized as either a single intersection, multi-intersection or a self-intersection. There are two specific cases of the intersection case which are *cross* and *touch*. Figure V.4 illustrates different single *touch* relationships between an

input road segment and a blade road segment. Regarding the cases illustrated by Figure V.4(a) and Figure V.4(b), the input road segments cannot be segmented by the specific blade road segments that *touch* the input road segment by a node of the input road segment. Figure V.4 illustrates the different cases that the blade road segment multi-intersects the input road segment. For example Figure V.5(b) shows a multiple *cross*, while Figure V.5(c) exhibits a multiple *touch*, and Figure V.5(a) a combination of *cross* and *touch*. The input road segments cannot be segmented by one of the intersections between an input road segment and a blade road segment where the blade *touches* the input road segment at a node of the input road segment.

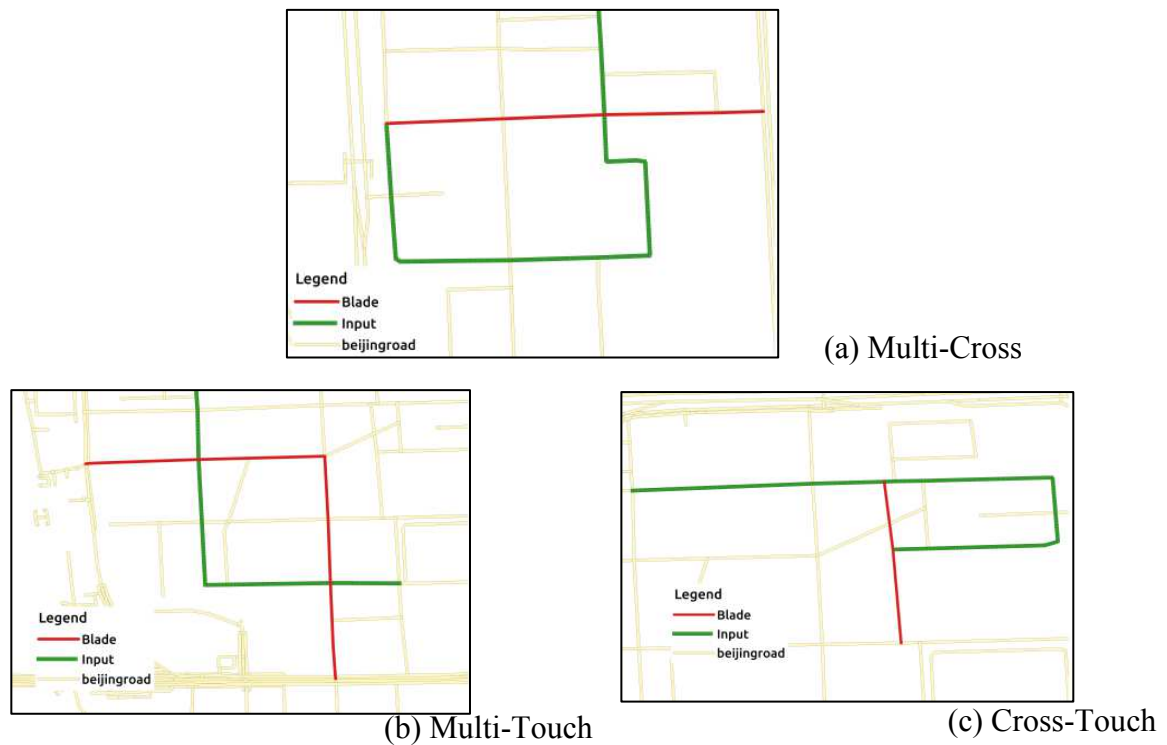


Figure V. 5. Multi-Intersection

In order to ensure the accuracy of the segmentation and network construction, all these situations discussed above are considered in the implementation. The design of the segmentation algorithm is shown in Figure V.6 and the overall implementation specified by Algorithm 2: The set of nodes is generated from the starting, ending and intersection nodes generated from the transportation road segments of the transportation network. The topological network is then built from

these road segment nodes, road segments, and road segment nodes to segment relationships. The implementation of the algorithm is performed on top of PostgreSQL/PostGIS database system using PL/PGSQL, a loadable procedural language of PostgreSQL. In order to evaluate the performance of the algorithm, transportation polylines within a random region selected in Beijing City are extracted and considered as a sample. The PL/PGSQL function implemented on top of the segmentation algorithm is tested in this sample region. The segmentation results show that all the transportation polylines across this region are appropriately derived as topological network. The evaluation on the algorithm applies a series of PostGIS functions such as *st_cross*, *st_touch* and *st_intersects* along with QGIS visualisation functions to check whether there are flaws in the segmentation process. QGIS is an open source GIS software that enables visualizations of raster and vector GIS data layers from PostGIS and GIS standards such as OpenGIS and Shapefile source files. It turns out that no mistakes can be visually identified, this being a very satisfactory result. Overall, 3277 transportation polylines crossing the sample region have been considered as inputs by the implementation. After the segmentation and network constructions thanks to the application of Algorithm 1 and Algorithm 2, 8591 *TransportationSegments* and 6180 *TransportationNodes* have been identified. The total length of the sample network is 700.252 Kilometres. Computing time consumption is 31.377 seconds which is largely acceptable for such process. Thus, one might conclude that Algorithm 2 is reliable with high accuracy and efficiency.

Algorithm 2 Segmentation

```
Begin    L ← a set of transportation segments    for every i ∈ L loop
        B ← set of candidate blades
        S ← set of segments
        set S = NULL
        if self-intersect(i) then
            {s1, s2} = split(i, intersection)
            {s1, s2} → S
        else
            i → S
        end if
        for every b ∈ B loop
            for every s ∈ S and intersect(s,b) loop
                if (touch(s,b) or crosses(s,b)) and !(touch(b, terminal of s))
then
                    split(s, b) → S
                end if
            end loop
        end loop
        S → result set
    end loop
return result set
end
```

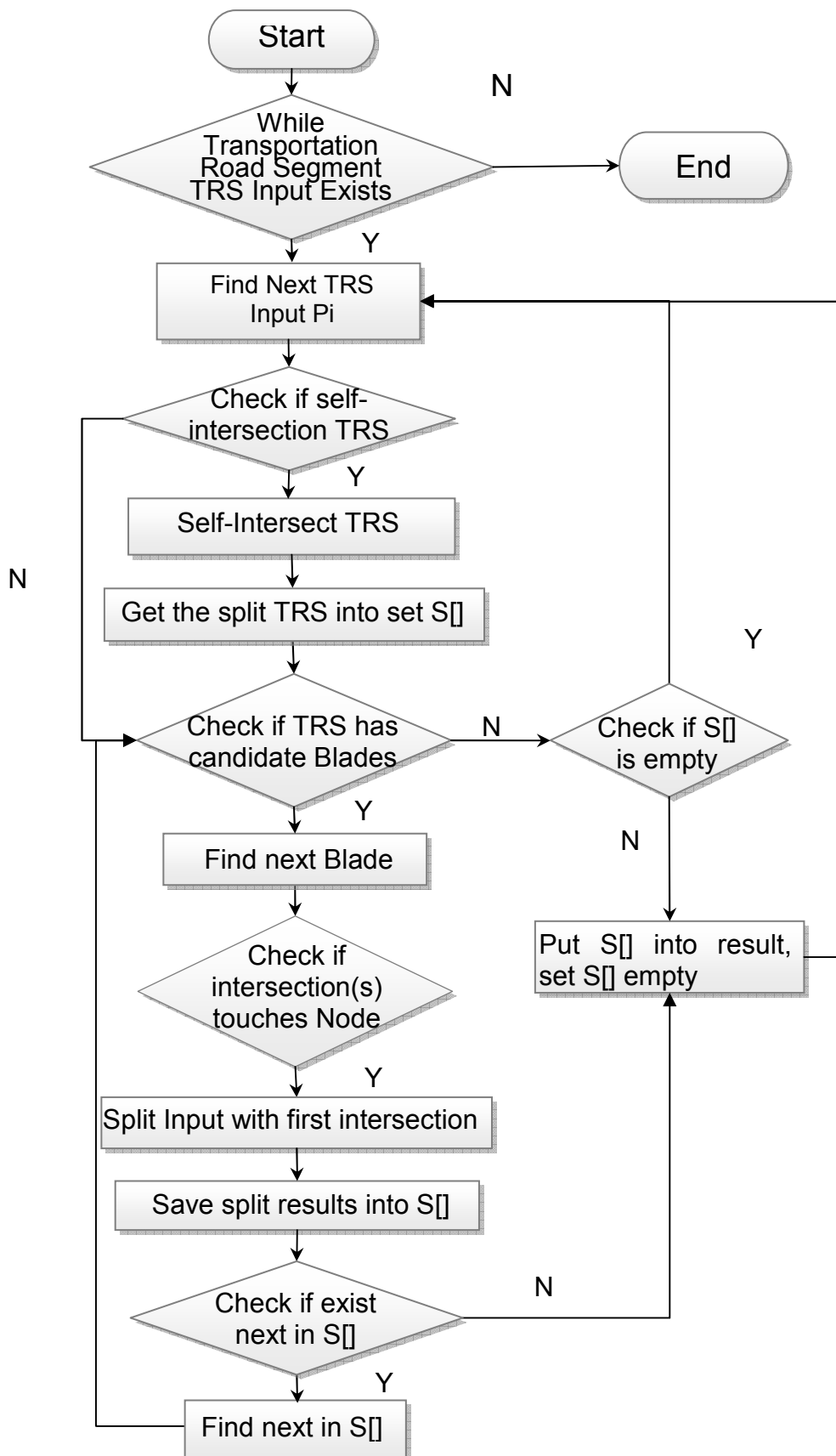


Figure V. 6. Algorithm 2. Transportation Network Construction

V.4.1.2 Map Matching

As mentioned in the previous sections, our travel displacement oriented research is based on the assumption that human beings preferably “move” along the transportation network when performing some activities in the city. However, when considering the available Beijing dataset, massive data gained from GPS records does not always stick to the spatial constraint due to some limitations of satellite positioning systems and quality of the available reference cartographical data (e.g., inaccurate GPS measurements, loss of smartphone signals, vectorization imprecision of scanned maps).

These measurement limitations are inevitable and part of many cartographical experimentations. Map matching has been first introduced by Honey et al. (1989) in order to match positions from a GPS dataset to a map. For network-constrained trajectories, map matching methods can be used at two reconstruction levels:

- (1) cleaning the trajectory by replacing each inaccurate moving object position on the trajectory by a point on the network that is the most likely position of that moving object;
- (2) transforming the raw trajectory into a semantic map-matched trajectory (Parent et al. 2013).

Several map matching algorithms have been so far developed (e.g., Brakatsoulas et al., 2005; Lou et al., 2009; Newson et al., 2009). From a generic point of view, map matching can be summarized by two cases: online and offline (Pereira et al., 2009). Online map matching mainly focuses on real-time applications such as GPS navigation. However, online algorithms are less effective but more efficient with low accuracy compared to offline map matching, as for the online algorithms concern more time consuming. Offline map matching rather operates on post-processing applications and then has better performance and will be retained.

An implementation of a map matching algorithm is indeed necessary to rectify the raw trajectory data and map it onto the transportation network, and this according to the principles defined by our modelling approach. The map matching algorithm implementation is mainly based on the topological method proposed by Brakatsoulas et al. (2005). In this approach (Figure V.7), for each point P_i in a trajectory, considering that P_{i-1} has been matched to a road segment, the adjacent segments to the segment are candidate segments to be matched to P_i . Considering the specificity of the original data and the one of the network constructed, we take

all the segments that might be in the path to the previous matched segment S_{i-1} within the total length lower than the possible distance the considered object of the trajectory might move.

Algorithm 3 Map Matching

Begin

$T \leftarrow$ a set of points in trajectory data

for every $p_i \in T$ loop

$D \leftarrow$ distance tolerant between the selected point to the candidate segment

$S \leftarrow$ set of segments

set $S = \text{NULL}$

for every $s_i \in (S \text{ in } \text{distance}(s, p_i) < D \text{ order by distance desc})$ loop

$l = \text{speed} * \text{interval}(p_i, p_{i-1})$ // possible distance travel at average speed within time interval

if $s_i = s(p_{i-1})$ then

$p_i = \text{closest_point}(p_i, s_i)$

else

if (s_i has path to $s(p_{i-1})$) and $\text{length}(\text{total path}) < l$ and $(\text{angle}(p_i, p_{i+1}) - \text{direction}(s_i)) < 90$ then

$p_i = \text{closest_point}(p_i, s_i)$

end if

end if

end loop

end loop

return T

end

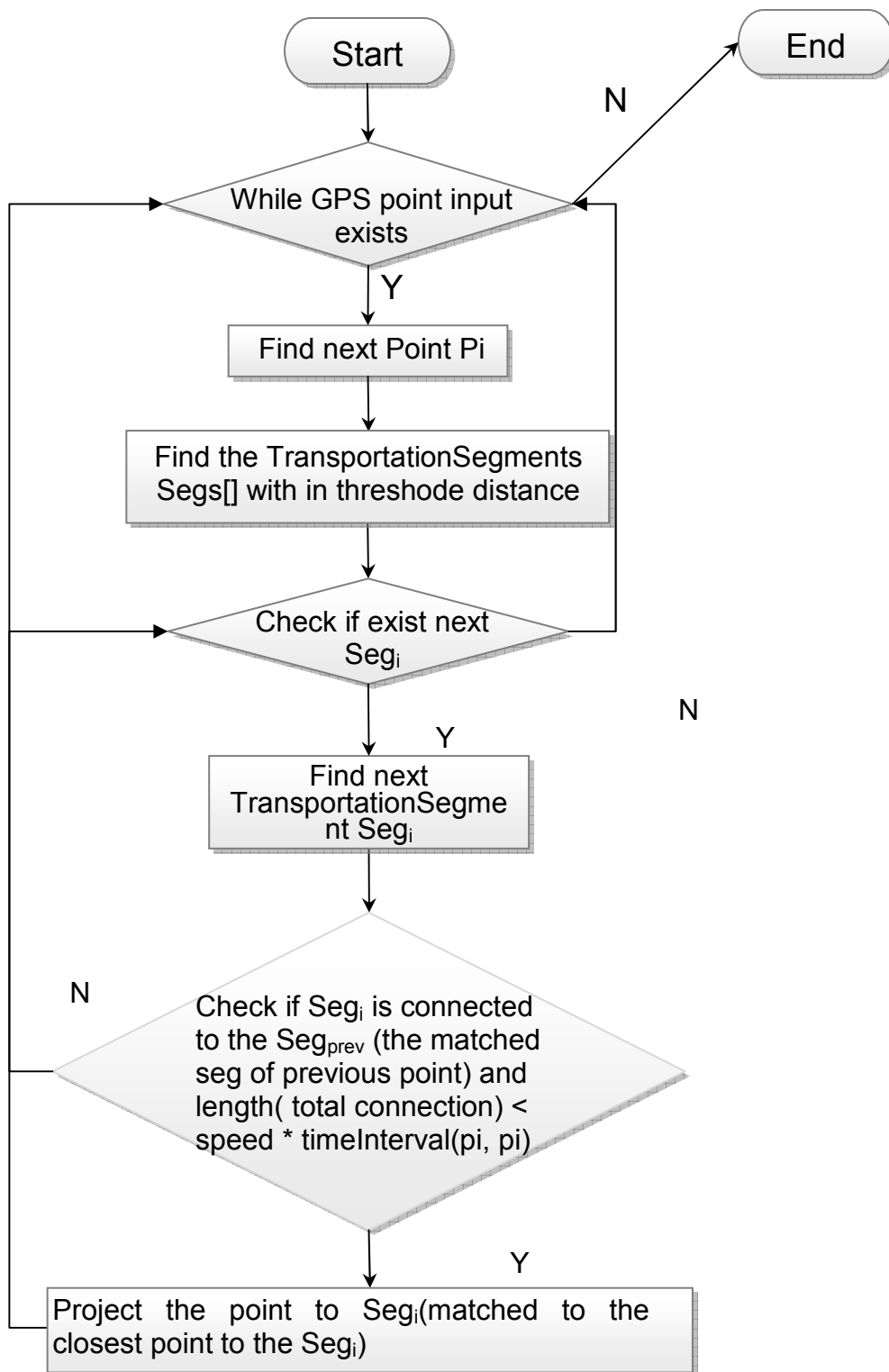


Figure V. 7. Algorithm 3. Map Matching

The algorithm considers three parameters to determine whether a trajectory point should be matched appropriately to a segment.

- Distance: The lower the distance between a candidate point and candidate segment, the higher priority is given to that candidate segment.
- Spatio-temporal consistency: connectivity with previously matched segments and travel distance are combined to refine possible segment candidates. Due to the density of the transportation network data of Beijing city, segments in the network are usually short. So not only adjacent segments should be considered, but also the ones that have lengths in line with realistic travel distances.
- Bearing: The travel direction is also considered as a validating parameter. The less difference between the bearing of the candidate Transportation Segment and that of the trajectory, the more likely they should be matched.

V.4.2 Database Construction and Analysis

The data pre-processing introduced in the previous section has provided a robust foundation for further data organization and analysis. This section develops the generation of the experimental database (section 4.2.1) and the principles applied for the routes extraction (section 4.2.2).

V.4.2.1 Data Structure Generation

The data structure is implemented and populated by the application of the algorithms introduced in the previous section. Trajectories embedded in the Geolife data sets constitute the input of the algorithms, while the output is organized in the form of the trajectories identified by our modelling approach. According to the Trajectory module developed in Chapter III, the transportation network is modelled by `TransportationSegments` and `TransportationNode` and populated accordingly. Moreover, temporal data is kept and represented as an additional temporal attribute associated to the map patching process. Every node part of a given trajectory should be modelled as such and time-stamped. This implies to algorithmically derive this information (Algorithm 4). The principles of the trajectory derivation function are demonstrated in Figure V.8. Along with the step `Insert`, the average speed of the trajectory in the matched `TransportationSegment` is calculated according to the distance between the two nearest points in the trajectory and the time interval

between them. Each segment node is time-stamped according to the speed associated to the nearest GPS point and the distance between them.

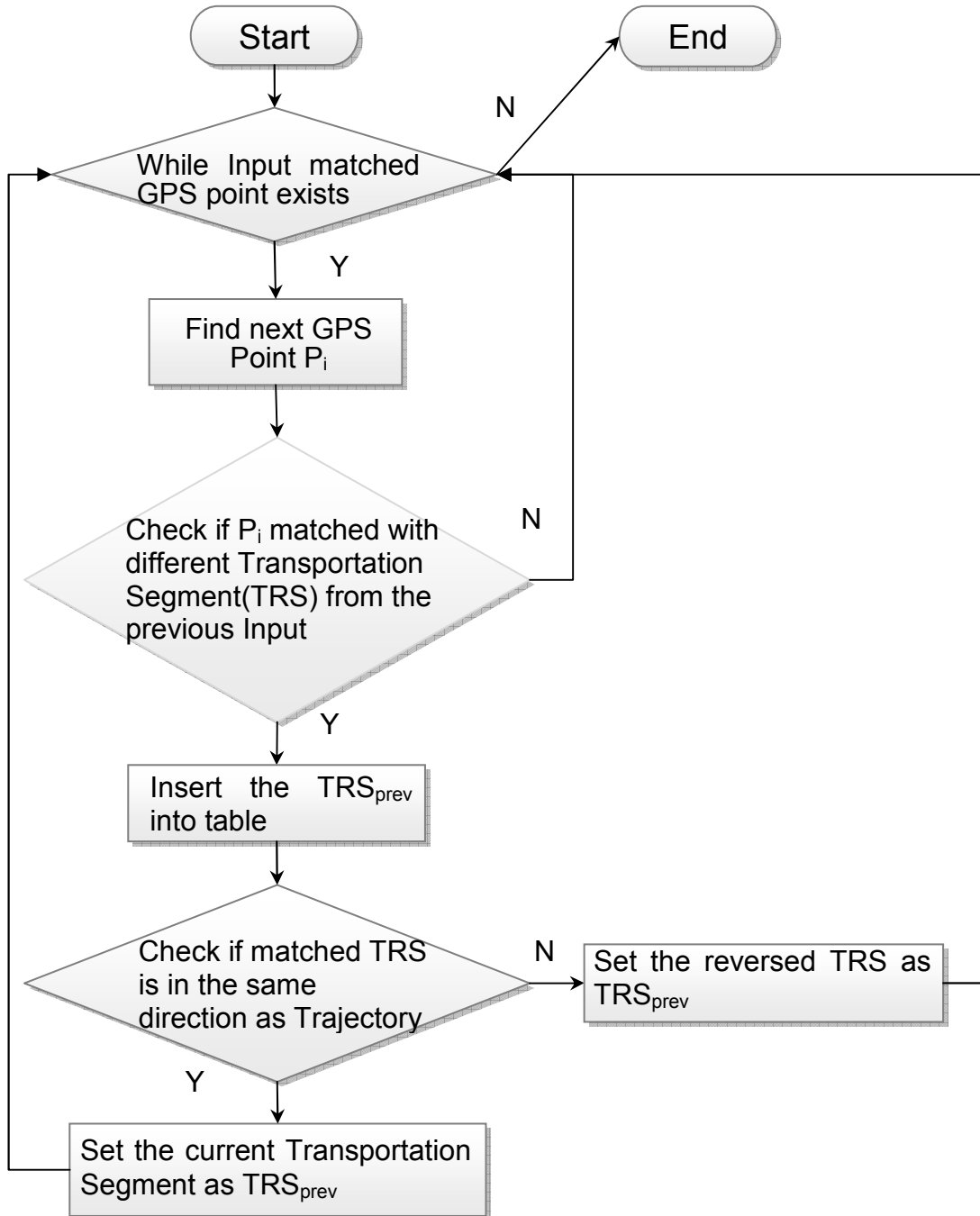


Figure V. 8. Algorithm 4. Trajectory Construction Principles

Algorithm 4 Trajectory Construction

Begin

$T \leftarrow$ a set of points from the map matching process

StartPT \leftarrow Start point of a trajectory segment

PrevSe \leftarrow TransportationSegment ID of the Previous Point's matching

for every $p_i \in T$ loop

 if $i = 1$ then

 set StartPT = p_i

 PrevSe = p_i .Se

 end if

 if p_i .Se \neq PrevSe then

 if co-direction(trajectory, PrevSe) then

 reverse the direction of PrevSe

 end if

 calculate the speed of the End of PrevSe

 EndPt.instant = p_{i-1} .time + Distance(p_{i-1} , PrevSe.EndPT)/speed of p_{i-1}

1

 StartPt.instant = p_{i-1} .time - Distance(p_{i-1} , PrevSe.StartPT)/speed of p_{i-1}

 insert into TrajectoryPartTable values(StartPT, the End of the PrevSe)

 set StartPT = p_i .Se.StartPT

 calculate the speed of StartPT

 PrevSe = p_i .Se

 end if

 if p_i is the last p_i then

 insert into TrajectoryPartTable values (StartPT, p_i)

 end if

end loop

return

end

V.4.2.2 Route Extraction Algorithm

According to our Travel Displacement Model, Route is a key modelling abstraction which contains both spatial and temporal attributes connecting the three modules: Trajectory Module, Transportation Network Module and Travel Pattern

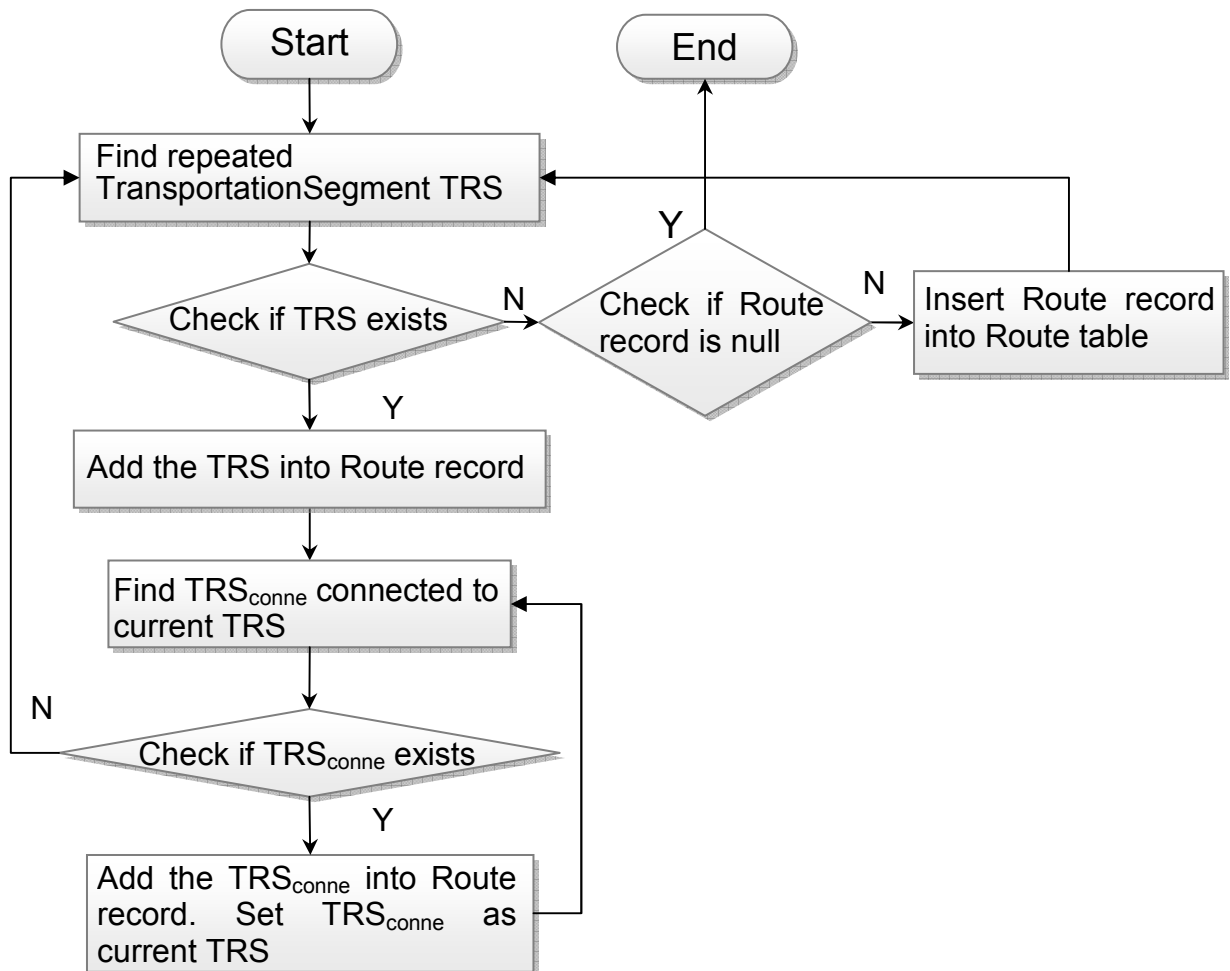


Figure V. 9. Algorithm 5. Route Extraction Algorithm

Module. A Route is defined as a sequence of regular spatio-temporal segments extracted from Trajectories. The route extraction algorithm considers not only regularly performed road segments, but also the consistency of these road segments. Therefore, this route extraction algorithm should extract consistent spatio-temporal *LineStrings* regularly taken by an individual. The algorithm is executed as follows(Figure V.9). First, all repeated *TransportationSegments* performed by an individual are identified. Next, the ones that are spatially and temporally connected are identified and sequentially ordered. The search continues until there is no more segments *TransportationSegments* are identified and spatially temporally connected.

Algorithm 5 Route Extraction

```
Begin
    SegmentList -> List of spatio-temporal Route Segments
    Select Segment From (Trajectories Segments that belong to a given individual)
    where Segment appears more than once into Segs
    for every Segment1 in Segs loop
        set PrevSe = Segment1
        SegmentList.append(Segment1)
        for loop
            find segment in Segs spatially and temporally connected with
PrevSe
            if segment is null then
                break
            else
                SegmentList.append(segment)
                set PrevSe = segment
                delete segment from Segs
            end if
        end loop
        insert into RouteTable values(SegmentList)
    end loop
end
```

These algorithms introduced in this section are essential data manipulations which organize raw dataset into well-formed spatial data structures as described in by the travel displacement model. The transportation segmentation algorithm plays a significant role in transportation network construction and populates the transportation module as defined by the travel displacement model. The map matching algorithm eliminates raw data flaws and matches GPS data into the transportation network, this enabling the trajectory module implementation. The trajectory construction algorithm populates the travel displacement module, while

the route extraction algorithm populates the Route object types. Overall, These algorithms provide a computational step-by-step approach to realize an implementation of the travel displacement model from raw trajectory datasets.

V.5 RESULTS AND DISCUSSION

An important objective of this experimental implementation is to not only to validate the applicability of our modelling approach, but also to evaluate its potential regarding the exploration of behavioural patterns. This section illustrates the data manipulation capabilities of the model by a series of queries that has been evaluated and performed on top of the Geolife dataset. This dataset has been populated in a PostgreSQL/PostGIS database and application of algorithms 1 to 4 as introduced in the previous sections. The data analysis potential of the whole approach is also illustrated by the resulting semantics extracted by this series of queries and whose complexity is variable. The main idea behind the application of these queries was to explore to which degree our modelling approach, when combined with advanced and built-in functions can extract some valuable patterns and outliers when analysing a very large trajectories datasets in the city.

Query 1: What are the road segments that Li Lei performed during the time interval 7am to 10 am? And what are the road segments that Li Lei regularly repeated during the time interval 6pm to 9pm?

```
Select Ind_find_route(id, '7:00:00', '10:00:00') from IndividualTable where name = 'Li Lei' ;
```

```
Select Ind_find_route(id, '18:00:00', '21:00:00') from IndividualTable where name = 'Li Lei';
```


The function `Ind_Find_Route(IndividualId, StartTime, EndTime)` is encapsulated in these two queries and whose results are illustrated by Figure V.10. This function returns the Routes followed by Li Lei during the time intervals given. These two queries implicitly make a difference between the routes performed by a given Individual early in the morning and late afternoon. It appears from the query results illustrated in Figure V.10 that the routes performed in the late afternoon are much more diverse than the ones performed in the early morning. This might reflect a relatively large set of activities performed by this individual after work, and on the other hand much less flexibility in the morning and more regular route and trajectory patterns. This is an illustrative first example on how the exploration of trajectory patterns can implicitly reveal some human displacement patterns in the city.

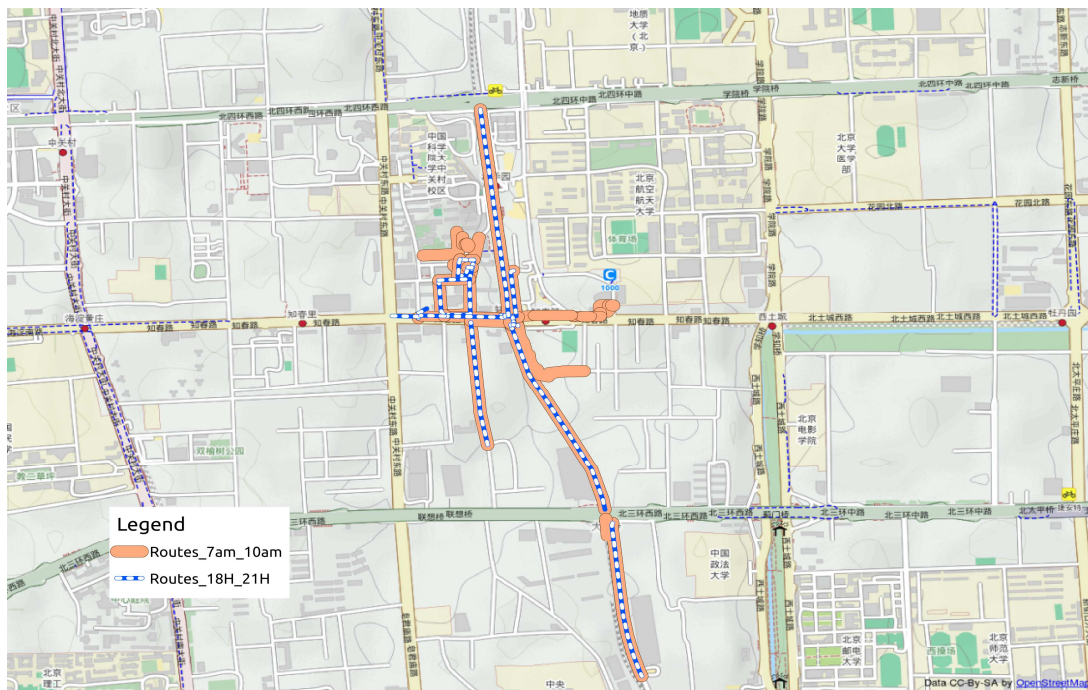


Figure V. 10. Transportation segments regularly repeated by Li Lei in the morning and evening

Query 2: What are the segments that Li Lei regularly repeated during weekdays and weekends?

Select `Ind_find_route(id, 'weekday')` from IndividualTable where name = 'Li Lei' ;

Select `Ind_find_route(id, 'weekend')` from IndividualTable where name = 'Li Lei';

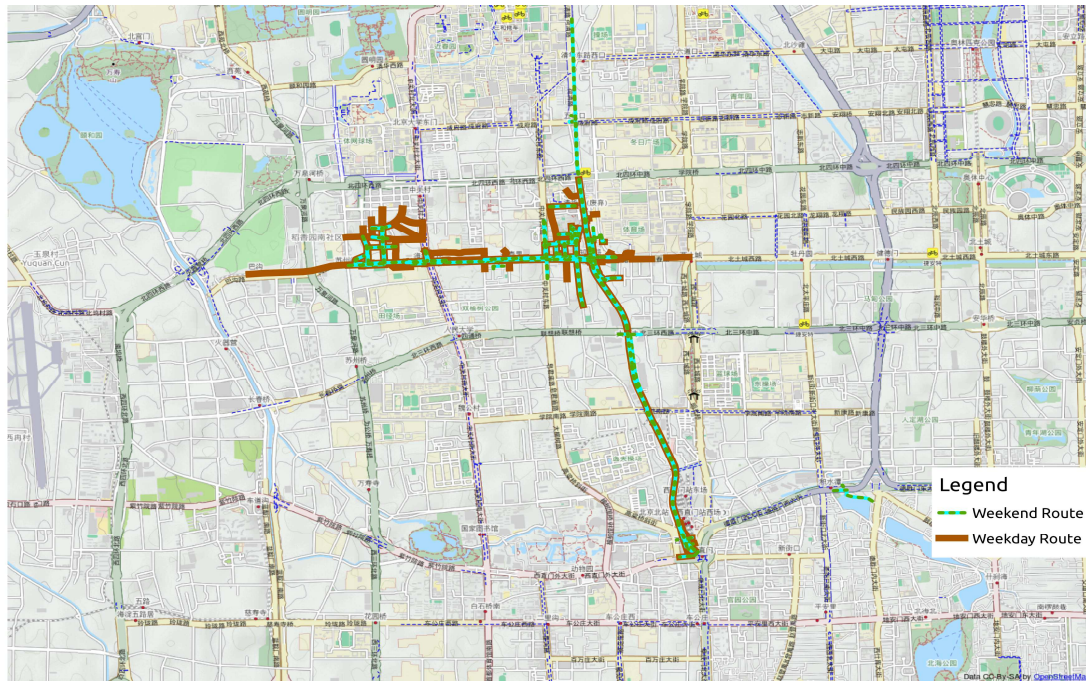


Figure V. 11. Weekday versus weekend routes for a given individual

The function *Ind_find_route* is applied in these two queries and searches for the routes performed by this Individual during weekdays and weekends. They allow to make a difference between people's habits in weekdays and weekends. Figure V.11 illustrates the range of routes follows. It clearly appears that the geographical range of the routes followed during weekdays is smaller than the range of the ones followed during weekends, this also shows regular activity patterns during weekdays and on the contrary a higher diversity of displacements during weekends with a larger geographical extent.

Query 3: What is travel extent covered by Wang Hua and Li Lei from 9am on Sept. 21 2008?

```
Select Traj_MBR(union(trajjectory.id))from TrajectoryTable traj, IndividualTable
indiv where traj.individualid = indiv.id and (indv.name = 'Wang Hua' or indiv.name
= 'Li Lei') and extract(hour from starttime)>9;
```

The function *Traj_MBR*(TrajectoryId) is encapsulated in this query. This function returns the minimum bounding rectangle (MBR) of the given trajectories.

The objective of this query is to reveal the respective geographical extent of the trajectories performed by some given Individuals for a given time interval. Rather than comparing the exact geometry of the trajectories performed by this given Individual, the idea behind this function is to visually compare the overall regions covered, these regions being derived by a MBR. When some given Individuals share some geographical regions, this might reveal some noteworthy patterns. Figure V.12 shows the result of this query applied to two given Individuals. In this specific case the figures that appear show that the two individuals share a close location at the end of their trajectories in the morning.

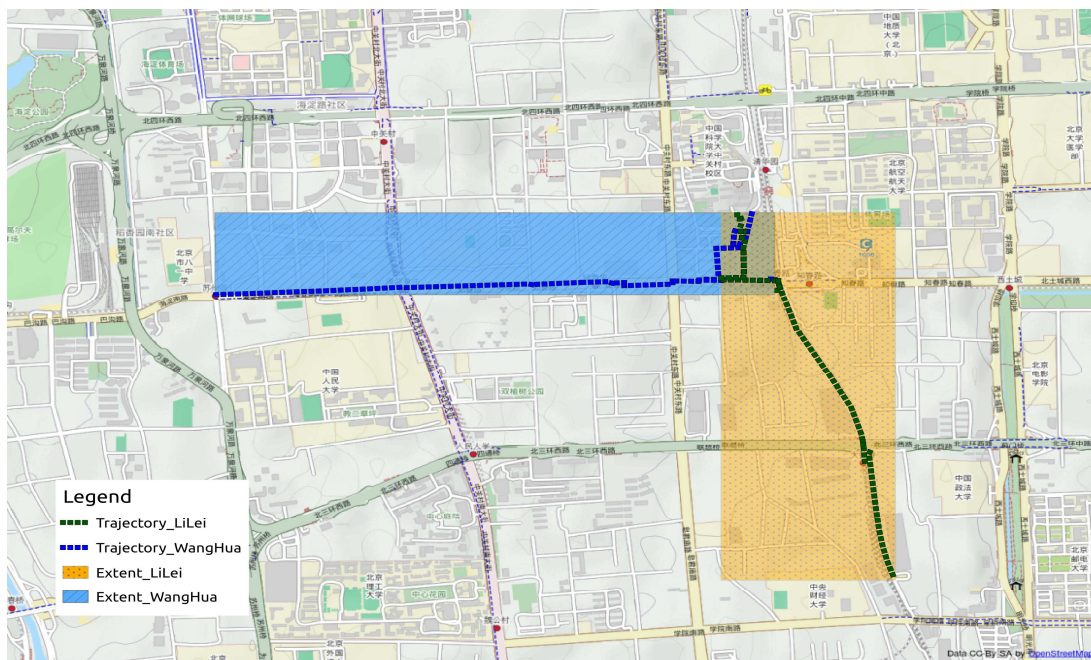


Figure V. 12. Geographical range of trajectories performed by individuals during a given time

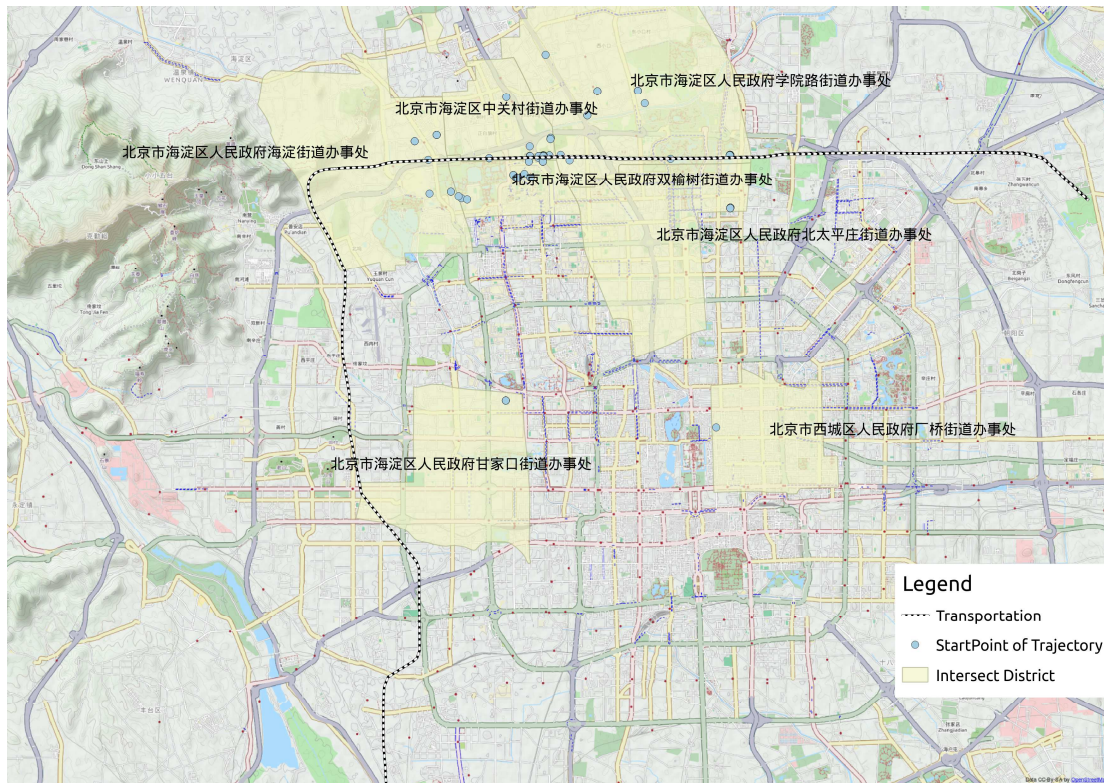


Figure V.13. Districts revealed by the start of the subway Line 10 trajectories

Query 4: Search for the districts some individuals that take the subway line 10 between 6am and 9am come from?

```
Select district.id, district.geom from district, TrajectoryTable trj,
TransportationSegmentTable seg, TransportationTable trans where
Traj_PassTransport (trj.id, seg.id) and extract(hour from trj.starttime) between 6am
and 9am and seg.roadid = trans.id and trans.name = 'subway line 10' and
st_within(trj.startpt,district.geom);
```

The function `Traj_PassTransport(Trajectory.Id, TransportationSegment.Id)` is applied to this query. This Boolean function returns *True* when a given *Trajectory* passes through a given *TransportationSegment*. The aim of this query is to explore and search for the locations people came from when taking the subway Line 10 in the morning. In the city of Beijing, this might allow to search for transportation crowd patterns, a potential valuable information for transportation planners. The result of that query is shown in Figure V.13. It appears that the *Individuals* who take

the subway Line 10 in the morning mainly came from locations in the North West of the city.

Query 5: Search for the most regular routes followed by individuals in the afternoon from 5pm to 7pm. And which are overall the most popular paths followed by most individuals in the afternoon from 5pm to 7pm?

```
Select rou.id, num, trans.name, trans.type, seg.geom from (select sum(array_length(trajs,1)) num, rou.segmentid from RouteTable rou group by segmentid order by num desc limit 15), TransportationTable trans, TransportationSegmentTable seg where rou.segmentid = seg.id and trans.id = seg.roadid;
```

```
Select rou.id, num_indiv, trans.name, trans.type, seg.geom from (select count(rou.individualid) indiv, rou.segmentid from RouteTable rou group by segmentid order by indiv desc limit 15), TransportationTable trans, TransportationSegmentTable seg where rou.segmentid = seg.id and trans.id = seg.roadid;
```

These two queries apply spatio-temporal aggregations. The results are illustrated in Figures 14 and 15. The aim of the queries is to find the most popular routes performed in the city for a given time interval as well as the busiest ones. This should reveal some aggregated travel patterns. The query results reveal the routes performed most regularly by individuals (first query, i.e., the top 15 routes) and the busiest ones (second query). Figure V.14 shows the busiest routes while Figure V.15 outlines the ones performed most regularly and might reveal some regular patterns when observing people's routes. It appears that the most popular routes are relatively dispersed in the city, while routes repeated by most individuals are more concentrated in one region. The number of aggregated individuals in the Figure V.15 is relatively small as a few people follow regular routes. This reflects the diversity of travel patterns in this city.

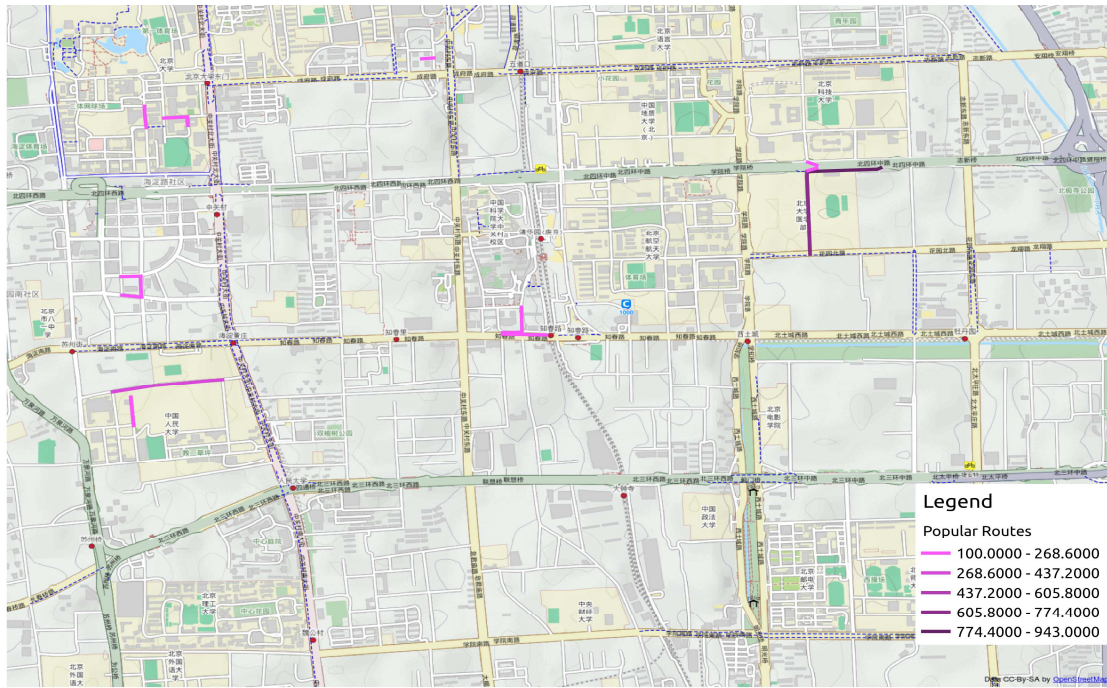


Figure V. 14 Busiest routes

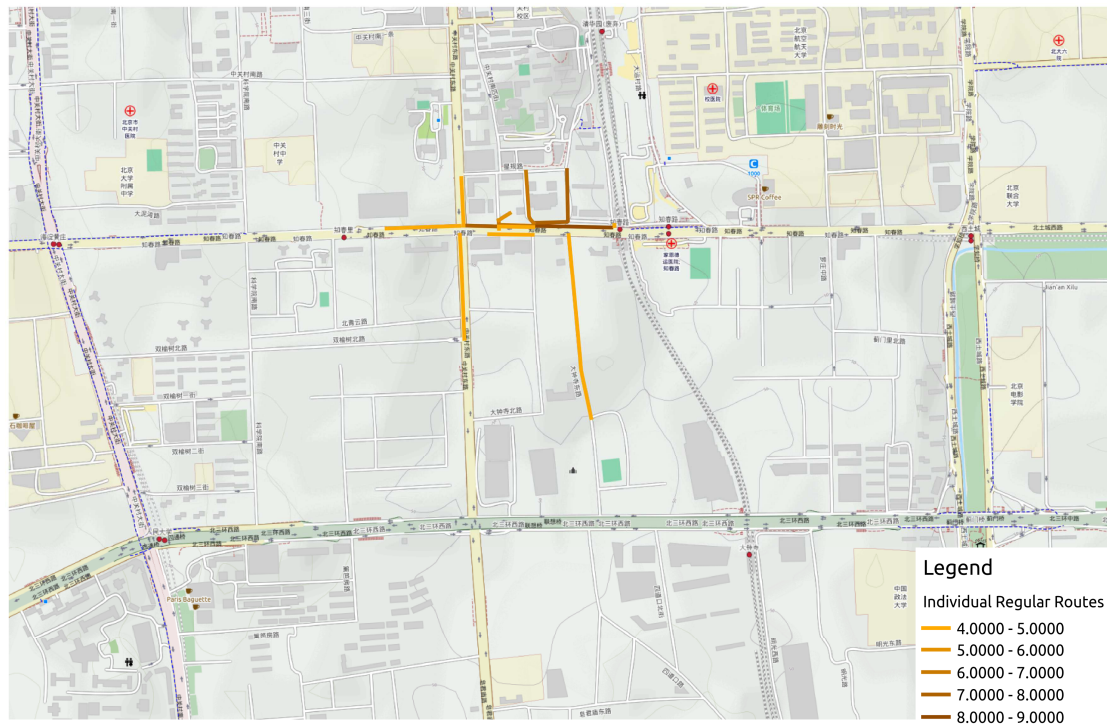


Figure V. 15 Most popular individual routes

Query 6: Where does the individual named ‘Li Ming’ usually start and end its trajectories before 9:00 in the morning? What are the possible destinations?

```
Select Traj_Nearest_POIs((select endpt from TrajectoryTable traj, IndividualTable indiv where indiv.id = traj.individualid and indiv.name = ‘Li Ming’ and extract(hour from traj.starttime) < 9),50);
```

The function `Traj_Nearest_POIs(Trajectory.StartPT/EndPT, Radium)` is applied to this query. This function returns the nearest Point of Interests near a Start or End of the Trajectory within the given distance. The query results illustrated in Figure V.16 show that morning destinations are relatively restrained to a small region. On the other hand trajectories performed in the morning reflect a relative diversity. As for the previous queries, it appears that in most cases the straight analysis of trajectory patterns can reveal some valuable displacement patterns, and this with almost no incoming data and information on the properties exhibited by the incoming trajectory dataset considered for this experimental evaluation.

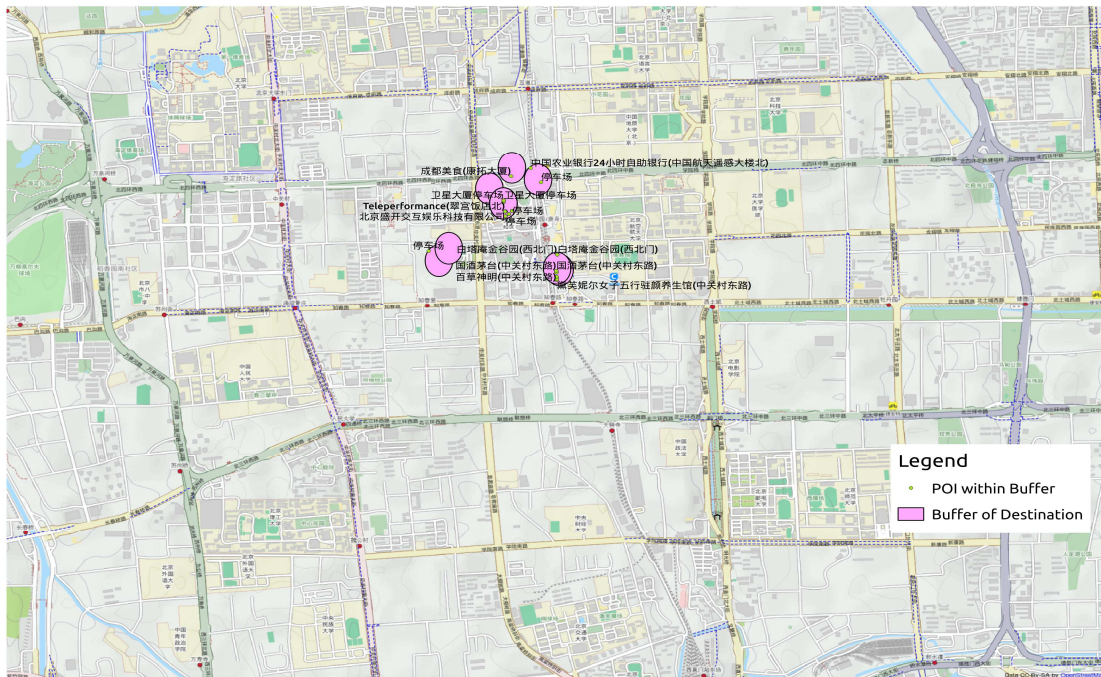


Figure V. 16. POIs near the destination of some trajectories in the morning

Query 7: “Find the range of the trajectories derived from the Beihang University in the afternoon.”

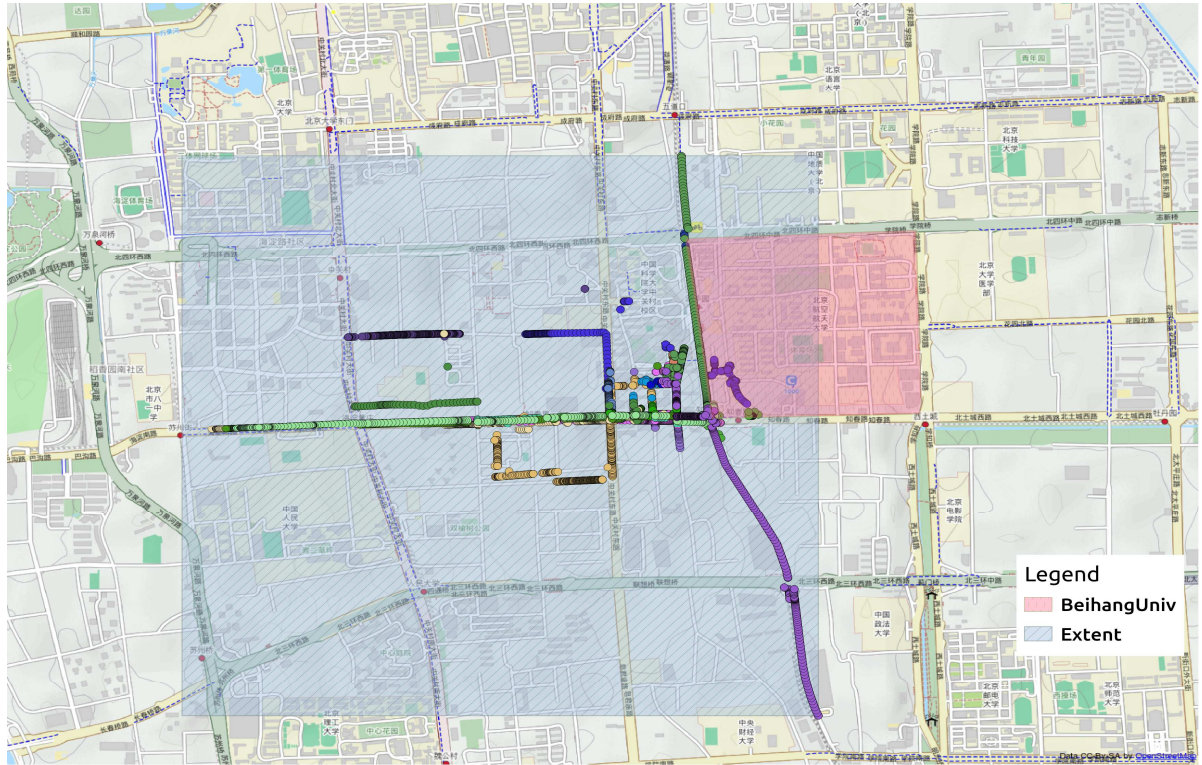


Figure V. 17. The Travel Range Derived from Beihang University (Region marked as light blue, the points are the trajectories moving from the university)

```
select Traj_MBR(select array(select traj.id from TrajectoryTable traj, LandUse land
where land.name = 'Beihang University' and st_within(startpt, land.geom) and
extract(hour from starttime) between 13 and 18;
```

This query searches for the range of trajectories derived from a certain POI (Point of Interest) or a certain area, in this case a university. By finding out the range of the travel oriented from a given place, one might derive the moving trend of peoples who leave the place. By finding out the range of trajectories (Figure V.17) that start from the Beihang University in the afternoon, one can also tell that people that leave the university usually tend to move in a South-West direction.

Query 8: “How many times did 'Li Ming' passed the road “三环” in the morning, in the afternoon and in the evening individually?”

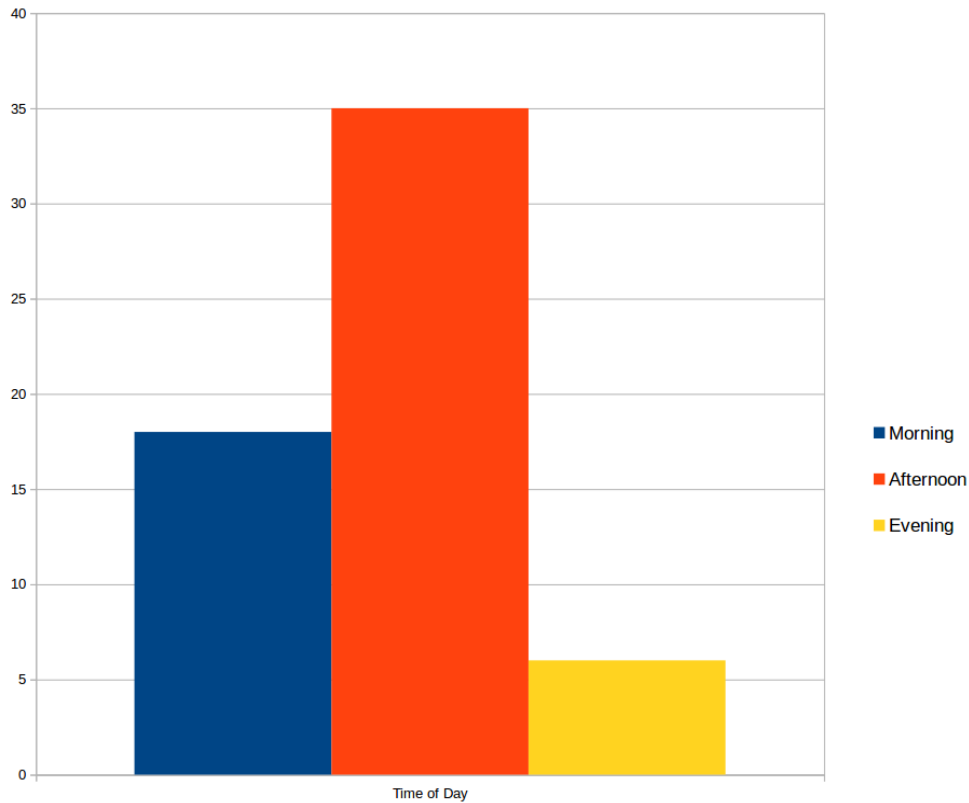


Figure V. 18. Li Ming's Trajectories passed by Road '三环'

```
Select trajectory.id, starttime, endtime from TrajectoryTable traj,
TransportationSegmentTable trans where trans.name = '三环' and tran.type = 'road'
and Traj_pass_transport(traj.id, trans.geom) and extract(hour from starttime)
between 7 to 12
```

This query digs the travel habits of a given individual. By finding out the frequent paths of a given individual at different times for a given day may help to search for a better understanding of travel displacements in relation with different traffic conditions. Figure V.18 shows that this individual passed through the road called ‘三环’ most frequently in the afternoon, and least frequently in the evening.

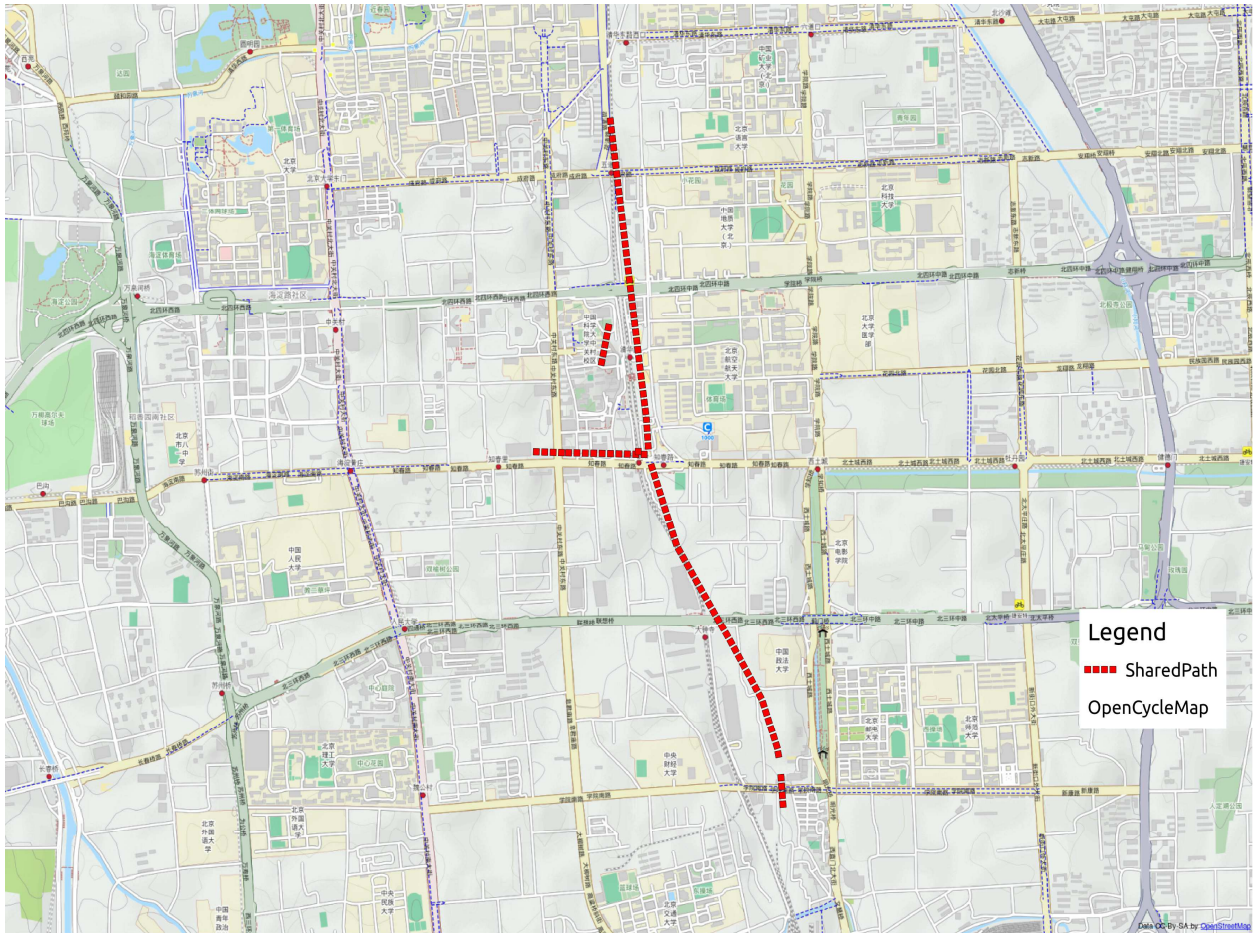


Figure V. 19. Shared Paths of Li Ming and Wang Lei

Query 9:“Find the paths that Li Ming and Wang Lei have shared.”

Select Ind_Share_Path(array(a.id,b.id)) from IndividualTable a, IndividualTable b where a.name = 'Li Lei' and b.name = 'Wang Lei'

This query searches for travel routine similarities of for two given individuals. The objective is to search for travel habits shared by two individuals. The Path is defined as an object type according to the travel displacement model and that models a sequences of transportation segments that a given individual usually follows. In fact this query figures out overlapped travel routes between two given individuals (Figure V.19). The potential of this category of query is to search for some individual routine similarities in order to develop collaborative transportation sharing.

This chapter reports on the implementation of the travel displacement framework. From the experiments developed it appears that the geographical database system PostgreSQL/PostGIS guarantees an efficient database representation and manipulation of a large raw trajectory dataset according to the principles of our travel behavior model. Moreover, the encapsulated functions available in PostgreSQL/PostGIS (e.g, extensibility, indexes) facilitate the whole implementation and the definition of many of the built-in functions introduced by our modelling approach. Most of the data manipulation functions suggested have been implemented as a proof-of-concept. The different query samples presented illustrate the potential of the approach according to the spatial, temporal and semantic levels. Overall, the implementation has been performed with a large raw trajectory dataset derived from GPS data of Beijing City. The results of the experimented have been thoroughly discussed, and show that our framework is capable of extracting individual travel mobility choices from an urban realm.



Conclusion & Perspectives

VI.1	OUTLINE.....	118
VI.2	MAIN CONTRIBUTION.....	119
VI.2.1	CONCEPTUAL MODELLING.....	119
VI.2.2	LOGICAL DESIGN	120
VI.2.3	DATA MANIPULATIONS & FUNCTIONS.....	121
VI.2.4	IMPLEMENTATION.....	121
VI.2.5	TRAVEL BEHAVIOUR EXPLORATION	122
VI.3	PERSPECTIVES.....	123
VI.3.1	MODELLING EXTENSIONS.....	123
VI.3.2	DATA ANALYSIS EXTENSIONS	124
VI.3.3	DATA MANIPULATION EXTENSIONS	124
VI.3.4	DECISION MAKING & TRAVEL PREDICTION.....	126

VI.1 OUTLINE

This PhD dissertation proposes a generic modelling framework for a semantic and spatio-temporal representation of human travel displacements in an urban environment. The framework includes a comprehensive travel displacement model aiming at favouring a better understanding of travel patterns. On top of the conceptual and logical models developed, we introduced a series of data analysis functions that have been implemented within a spatio-temporal query language as shown in Figure VI.1. The whole approach favours the exploration of citizens' travel patterns. The potential of the framework has been illustrated by an application to the Geolife reference database available in the city of Beijing. One of the interest of our approach is that it shows the feasibility of a spatio-temporal database integration of a large trajectory dataset, as well as the development of query capabilities that favour the exploration of mobility patterns in the city. This chapter first outlines the main findings and limitations of our research and then discusses possible methodological and research extensions.

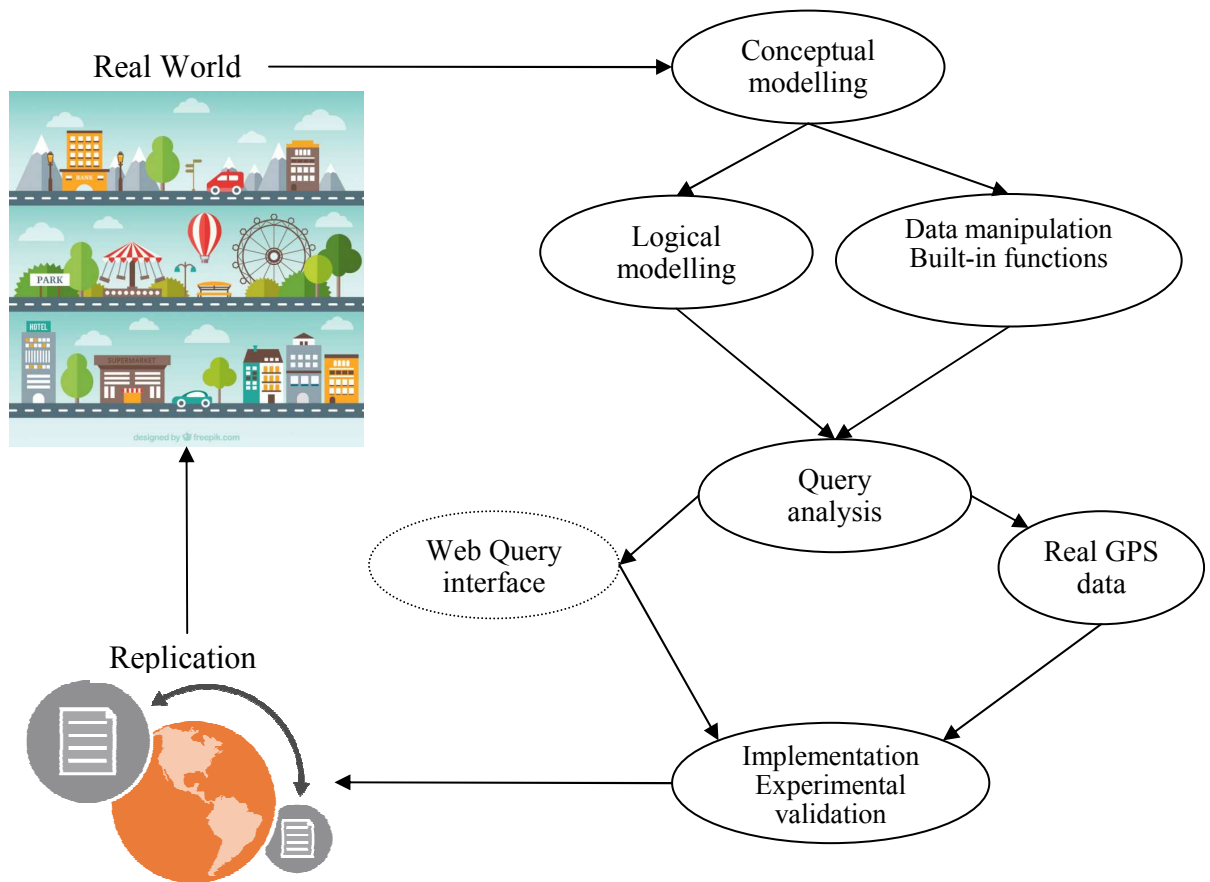


Figure VI. 1. PhD Research approach

VI.2 MAIN CONTRIBUTION

The contribution of this research includes several methodological and computing developments that respectively cover the design and implementation of a travel displacement database framework, and the exploration of human's travel patterns in a given city. As such our research can be considered at the crossroad of geographical information science and spatio-temporal databases as well as offering some application perspectives for urban studies.

VI.2.1 Conceptual Modelling

The first contribution of our research is a conceptual data model oriented to the representation of human travel displacements in a city realm. This approach is

designed and developed on top of the MADS model that presents the advantage of encapsulating several semantic, spatial and temporal modelling abstractions that provide a strong support for the development of a domain oriented database schema. We introduce a model in which spatial, spatio-temporal and semantic objects types are identified as basic modelling abstractions for the representation of citizen's travel displacements. In particular, these object types include several entities such as *Trajectory*, *Individual*, *TransportationSegment*, *TransportationNode*, *Route* and additional constraints that influence travel displacements. These object types are semantically represented and illustrated when necessary; object attributes are listed and defined according to the spatial, temporal and semantic dimensions. In order to reflect the whole semantic of a given application, object relationships are included in the model as well as appropriate cardinalities. These relationships also consider the semantic, spatial and temporal levels. Most of the relationships identified allow for a qualification of spatio-temporal object type *Route* that plays a specific and central role in our model. The *Route* object, which is defined as a list of repeated trajectory segments, connects the three model modules as a whole modelling framework.

VI.2.2 Logical Design

Not only the conceptual model provides a synthetic view of a human trajectories and mobility patterns, it also provides a pathway towards a logical model and then further implementation on top of a spatio-temporal database. In fact the conceptual model of human travel displacements acts a foundation for exploring some specialized data manipulations and functions that favour exploration and discovery of travel patterns in a given city. The conceptual model is then mapped to a logical representation using an extended geographical and relational database. Object types and relationships are mapped towards geo-relational constructs as usually done in geo-relational database modelling. One of the advantages of the MADS background model is that it favours the mapping of all of our modelling constructs to the logical model, this facilitating the initialization of all geo-relational tables within the PostgreSQL/PostGIS software environment.

VI.2.3 Data Manipulations & Functions

One peculiar contribution of our research is that several specialized functions that clearly encapsulate some semantic and spatio-temporal knowledge have been identified, categorized and then implemented. This choice is not only considered as an extension of a conventional extensible geographical query language, but also as overall providing some built-in and specialized domain-oriented manipulation functions appropriate for a human travel displacement database. The series of functions developed cover the semantic and spatio-temporal functions, they are by no means complete but show how some typical transportation-based queries can be encapsulated in a sort of extended query language. One of the main advantages of these functions is to clearly simplifying the writing and coding of these queries, this being often necessary at the user interface level.

VI.2.4 Implementation

The travel displacement model and database as well as the series of functions have been implemented on top of the geographical database system PostgreSQL/PostGIS. The integrated and basic spatial data manipulations available in PostGIS enable the processing of most of not all the data manipulation queries identified, as well as the embedding of the functions developed as part of the query language. The dataset used for the case study implementation came from the Geolife project that offers a very large dataset of real GPS travel data of Beijing City. In order to provide a manipulate data reference, Geolife data has been combined with additional transportation and district data, points of interest derived from the Baidu repository. Several geometrical algorithms have been designed and implemented in order to match the trajectories available to the underlying urban network. The algorithms developed combined data pre-processing and topological matching of the Geolife trajectories towards the Beijing road network. Overall the algorithms developed have proven to efficiently generate a manipulable and accurate enough trajectory and urban network spatio-temporal database. All algorithms have been implemented within PostgreSQL/PostGIS using the PL/PGSQL language embedded in PostgreSQL. They allow generating the main *Trajectory-TransportationSegments* and regular *Routes*.

VI.2.5 Travel Behaviour Exploration

The database framework implemented is evaluated using a series of typical travel pattern discovery queries. These experimental queries combine spatial temporal and semantic queries and specialized functions. These queries cover different levels of abstraction, including manipulation of single trajectories of one individual, travel routines and habits of several individuals, spatio-temporal relationships between individuals when performing some trajectories, as well as some general traffic flow queries. Overall, the results of the experimental test queries reveal some specific individual travel patterns as well as some traffic conditions at specific times. Most of the data manipulation and query results are visualized and discussed. Without loss of generality the experimental evaluation shows that the implementation of our modelling approach is feasible and also provides some appropriate query facilities at the interface level. The query results also show that the implementation has proven to be robust and computationally efficient.

In conclusion, our approach provides a modelling and computational approach for representing travel trajectories from a very large dataset and then integrating the whole within an extensible geographical database system, these favouring further manipulations thanks to built-in semantic and spatio-temporal functions embedded at the query interface level. Through the implementation on top of a geographical database system, the approach has proven to be appropriate enough to extract human travel habits in an urban environment.

Although the research has achieved the goal of discovering some travel displacements in a city realm, there are still several drawbacks and limitations to our work. Even though the travel displacement model contains an exhaustive sets of abstractions identified from the semantic of the application, when matching the model to the logical and physical levels, still not all the concepts identified are populated as in our case the Geolife dataset is currently limited to spatio-temporal data with a lack of semantic information. In fact this denotes a sort of recurrent problem when applying modelling approaches to real data: there is often discordancy in between the expressivity capability of the domain-based model

identified, and the one of the available data. Therefore, several additional data has been considered at the implementation level: POIs, geographical data, this making the data manipulation not completely integrated at the logical and data manipulation levels.

VI.3 PERSPECTIVES

Despite the potential interest of our research framework, several research and application directions are still left to further work. These cover several dimensions from the conceptual to the logical models, computational and implementation issues, data analysis and decision-making as well as the exploration of different application domains.

VI.3.1 Modelling extensions

Since the design of the travel displacement model is based on the MADS modelling method that provides sufficient orthogonality between the different modelling dimensions, as well as in between the conceptual and the logical level, this provides many opportunities for further extensions without denaturing the main principles of the current modelling approach developed so far. MADS also has the advantage of achieving simplicity, and strong expressive power. This in fact really favours extensibility and integration of additional semantics. For instance, the semantic of the different travel displacements might be revisited and/or extended according to some specific constraints coming from the application level, this being often the case as our domain of study is relatively specific, and every urban travel database might own some very specific properties to be taken into account. In particular, causalities might be further studied and also integrated as part of the semantic model, this ensuring a better connection in between the patterns that potentially emerge from a large trajectory dataset, and some causal information available. This notion of causality is closely connected to the different behavioural and factorial dimensions that generate and constraint such human travel displacements in the city. Such extensions are likely to provide deeper understanding and comprehension of travel displacements.

VI.3.2 Data analysis extensions

The data analysis extensions can be closely associated to the travel patterns module as identified by our conceptual model. Although the current travel displacement model categorizes and defines the constraints that generate some travel patterns, there is still lack of substantial analysis and algorithm that might support further reasoning and exploration of travel patterns. In order to extract and identify how some specific factors influence travel displacements, further data analysis and mining processes are required in order to manipulate appropriately the spatio-temporal and semantic levels. Since such analysis should be mainly based on the data available, as well as on its intrinsic quality, additional trajectory datasets will be needed in order to revisit and generate some travel displacement databases. One might also consider integrating additional datasets such as weather conditions, activity-based data, personal profiles etc. Overall, the larger the input databases, the larger the potential of data exploration will be offered at the data manipulation level to reveal and explore travel habits in the city.

VI.3.3 Data Manipulation Extensions

Assuming that additional and richer input datasets will be available under the principles given in the previous section, there will be a need to develop additional algorithms to take full advantage of them. Additional query mechanisms will be designed in order to search for the potential reasons that lead to some specific travel displacements according to the extended pattern module. In order to achieve that goal, Specific queries might for example search “How often individuals usually travel by bus instead of subway for a specific route?”. Or “When individuals are likely to change their routes?”. These examples show the range of queries that might be explored when giving a sound representation of human trajectories. Additional statistics on the semantic attributes available as closely related to travel patterns will be developed in order to complete the expressive power of the data queries. This reveals that in fact the data manipulation language (i.e., queries) should be considered as a path towards a better comprehension of travel habits, but additional capabilities should be used and statistical analysis being indeed one of them.

A direction still to consider in further work will be to make every effort to have better designed trajectory databases, as well as more robust integration of the geographical and trajectory databases. Still, and as shown by the series of queries implemented, current SQL-like language is often difficult to read and interpret for non-database specialists. There is surely a need to develop more user-friendly interfaces that might be used on top of our data manipulations. For instance and amongst many directions to explore the development of specific interfaces connecting the database level to a Web based map system is an appropriate one to explore. Such system will allow the users to custom their own queries according to some specific rules. Such developments can be realized with some popular systems such as MapServer.

Overall, and despite the interest of the experimental validation, we currently plan to extend our validation to other contexts as considering a dataset from a single city is not enough to guarantee the mining capabilities of our database framework. In order to evaluate the genericity of the modelling and computational approach proposed in this thesis, a trajectory dataset of the city of Shanghai city from taxi trajectories will be considered as a second case study still towards the exploration of human travel habits in the city.

Moreover, additional non-conventional data sources can be also considered as valuable data sources for exploring travel displacements in the city. Social networks provide for instance novel promising data sources where human beings report on real time some travel habits and thoughts. Although the Geolife database used for the case study was non-structured, many available urban data sources are much more structured this providing additional opportunities for exploring travel habits in the city, especially when combined with conventional approaches. Last but not least, and in order to combine the exploratory component of the principles behind data manipulation queries, data mining approaches will provide different but also inspiring technics for exploring for travel patterns in the city.

VI.3.4 Decision Making & Travel Prediction

The data manipulation and exploratory capability of the modelling and database framework proposed in this thesis can facilitate its application to many urban studies. One of those applications is to apply the framework to travel decisions. For instance, a travel decision assistant application can be built based on the travel displacement framework and the historical patterns available in the database. According to some travel conditions given by the user the system might match such constraints to previous travel patterns in the city in order to search for the most performing one according to some given conditions. Another possible application that this research might be of benefit is the one of traffic flow prediction in the city. By acknowledging the travel habits of the citizens in the city and their evolution, traffic conditions might be predicted at different levels of granularity.

Overall, these very few examples illustrate some of the possible directions to explore when studying travel patterns in the city. Indeed this modelling approach can also provide a sort of reference of completely different domains of study, but where the basic modelling dimension is the notion of trajectory. This might not only apply to human trajectories at large, but also to any application where the notion of trajectory is present, whatever the domain of study.



Résumé étendu de la thèse

A.1	INTRODUCTION	130
A.2	APPROCHE DE MODELISATION.....	133
A.2.1	MODELE DE MOBILITE.....	133
A.2.2	MODULE TRAJECTOIRE	134
A.2.3	MODULE TRANSPORTATION NETWORK	135
A.2.4	MODULE TRAVEL PATTERN	136
A.2.5	RELATIONS ENTRE DIFFERENTS MODULES	141
A.3	ANALYSE DE PATRONS DE DEPLACEMENTS	142
A.3.1	INTERROGATIONS DE COMPORTEMENTS DE DEPLACEMENTS (TRAVEL DISPLACEMENT QUERIES) 142	
A.4	IMPLEMENTATION ET VALIDATION	144
A.4.1	CONCEPTION DE FONCTIONS.....	145
A.4.2	DONNEES EXPERIMENTALES DE TRAJECTOIRES.....	147
A.4.3	EXTRACTION DES TRAJECTOIRES	148
A.4.4	RESULTATS	148
A.4.5	EVALUATION DES PERFORMANCES.....	157
A.5	CONCLUSION	158

A.1 INTRODUCTION

Les villes ont toujours été des environnements dynamiques et complexes. En effet, les villes sont le lieu d'un grand nombre d'activités humaines et d'interactions dont leur exploration et leur compréhension sont loin d'être aisées. En particulier, les grandes villes modernes génèrent de nombreux déplacements et des flux de trafic qui nécessitent encore le développement de cadres de modélisation de données spatio-temporelles appropriés afin de fournir aux planificateurs urbains des outils permettant une meilleure compréhension des patrons de déplacement.

L'étude des dynamiques de mobilité en milieu urbain est un sujet de recherche essentiellement abordé depuis le début des années 1980 (Hanson & Huff, 1986; Jiang et al, 2012; Pentland & Liu, 1999). L'une des principales raisons de l'émergence de ce courant de recherche est une conséquence du constat que la prédiction des comportements de déplacement des individus pourrait être un élément essentiel d'une meilleure planification des transports urbains et de l'analyse des politiques publiques de la ville (Kitamura, 1988). Un objectif important de la recherche d'une meilleure compréhension des dynamiques de mobilité passe par le développement d'une capacité de représentation et de prédiction des mobilités afin d'étudier comment en particulier les individus réagissent aux changements dans leur environnement. Cette recherche d'une meilleure compréhension des mobilités des individus reste encore une tâche difficile dans la mesure où cela implique non seulement des processus décisionnels complexes, mais également une série de contraintes sémantiques, spatiales et temporelles. A ce stade, le développement de méthodes et de modèles conceptuels d'exploration de données de mobilité pour une meilleure analyse et compréhension des comportements humains est toujours considéré comme un défi scientifique.

Plusieurs chercheurs ont successivement introduit des cadres méthodologiques et conceptuels pour représenter et analyser des comportements de mobilité (Alvares et al., 2007; Hanson & Huff, 1982; Fernandez & Yuan, 2000; Mennis & Peuquet, 2000; Yuan & Goodchild, 2007; Renso et al., 2013). D'après Andrienko *et al.* (2008), un mouvement est défini comme une généralisation de plusieurs primitives géométriques et sémantiques dérivées à partir de trajectoires individuelles, et qui peut être analysé à différents niveaux d'abstraction pour en dériver de la connaissance. A un niveau conceptuel, une trajectoire peut être représentée comme l'évolution de la position d'un objet en mouvement dans l'espace pour un intervalle

de temps donné et avec un objectif décrit par ses caractéristiques sémantiques (Spaccapietra et al., 2008). Au niveau applicatif, plusieurs taxonomies ont été introduites pour caractériser des patrons de trajectoires (Dodge et al. 2008).

Plusieurs algorithmes géométriques ont également été proposés pour décrire des mobilités. En particulier plusieurs approches de « clustering » de trajectoires ont été introduites par différents auteurs (cf. Nanni, 2013). Un patron de trajectoire sera défini par un ensemble de trajectoires qui recèlent des similarités spatiales, temporelles et sémantiques. Cette notion de similarité s'appréhende par exemple à partir d'une mesure de distance sémantique (Parent et al. 2013). Par ailleurs, Laube introduit plusieurs algorithmes de détection de patrons de mouvement à partir de la construction d'une matrice de mouvements et de contraintes de distance (Laube, 2005). Une approche plus spécifique appliquée à des mobilités urbaines a été introduite par Buchin et al. (2011). Cette approche est basée sur un calcul de la distance de Fréchet dérivée de deux trajectoires pour un même intervalle de temps donné, et ce, afin d'en dériver s'il y a lieu des patrons de similarité. Hung et Peng (2009) proposent une approche basée sur les probabilités et un algorithme de "clustering" également basé sur une fonction de distance. Une K-distance dérivée du K-voisin tel que défini par (Cover & Hart, 1967) peut également constituer la fondation d'un algorithme de « clustering » (Ong et al. 2010). Globalement, non seulement des patrons de mouvements peuvent être identifiés à partir de données de mobilités, mais aussi des trajectoires anormales ou particulières.

Les patrons de mouvement doivent idéalement révéler toute régularité spatiale et temporelle reconnaissable, sous réserve d'un intérêt qui dépend du domaine d'application (Dodge, 2008). Les irrégularités sont également des tendances à explorer. Par exemple, la découverte de déplacements aléatoires et particuliers au sein de larges bases de données de trajectoires est une autre direction qui a été explorée par Gonzalez et al. (2008). Dans de nombreux cas, les humains ont tendance à effectuer des déplacements routiniers, c'est un premier contexte d'exploration, un deuxième serait celui d'analyser des comportements particuliers ou même anormaux. Dans le cas des comportements réguliers, les contraintes institutionnelles, culturelles et obligations qui influencent les mobilités humaines jouent un rôle déterminant (Buliung et al. 2008, Schlich & Axhausen, 2003). Schlich a montré, qu'en général, l'étude de similarités dans les déplacements humains à travers de longues périodes peut révéler des tendances significatives (Schlich & Axhausen, 2003). De telles tendances émergent en général de l'étude de

déplacements journalier, voire de différences entre patrons de mobilité en semaine ou en week-ends. Ces aspects ont été étudiés dans de nombreuses villes canadiennes et européennes (Axhausen *et al.*, 2002; Buliung *et al.*, 2008; Susilo & Axhausen, 2007). D'autres études ont montré une relative instabilité de patrons de déplacements durant les fins de semaines (Srivastava & Schönfelder, 2003; Susilo & Kitamura, 2005).

Bien que de nombreux travaux antérieurs aient notamment montré l'existence de comportements de routine dans les mobilités urbaines, il reste encore nécessaire de développer des modèles à la fois sémantiques et spatio-temporels permettant d'intégrer des données de mobilités et les associer à une description fine de leurs caractéristiques. De telles approches devraient permettre non seulement de donner un cadre précis et manipulable de ces données, mais aussi de faciliter une plus large palette de manipulation et d'analyse à partir du moment où cet ensemble sera intégré au sein d'un SGBD extensible disposant de tous les outils de manipulation et de visualisation, et ce, à différents niveaux d'abstraction et d'échelle dans l'espace et le temps. L'objectif d'une telle approche consiste à pouvoir aborder des questionnements tels que l'étude des comportements habituels des usagers de la ville à différents niveaux d'abstractions dans l'espace et dans le temps, par quartier, par période temporelle, ou l'étude des modalités de déplacements et leur prise en compte par les réseaux de transport existants. Ces quelques exemples illustrent le type d'analyse qu'un tel modèle de base de données spatio-temporelle de trajectoires doit pouvoir aborder. Pour ce faire, nous proposons de combiner une approche de modélisation conceptuelle avec une démarche d'implémentation au sein d'un SGBD spatio-temporel extensible et une expérimentation appliquée à une large base de données de trajectoires en milieu urbain extraite du projet Geolife développé ces dernières années dans la ville de Beijing en Chine.

La suite de l'article est organisée en plusieurs sections. La section 2 développe l'approche de modélisation conceptuelle de représentation de mobilités urbaines à partir de plusieurs modules complémentaires qui permettent de caractériser les propriétés sémantiques, spatiales et temporelles. La section 3 propose une série de manipulation appliquées à ces trois dimensions et qui permettent d'illustrer le potentiel de l'approche. Une implémentation à partir du SGBD spatio-temporel Postgres/PostGIS et expérimentation aux données Geolife de la ville de Beijing est développée dans la section 4. L'article se termine par une conclusion dans la section

5 qui résume l'apport et quelques perspectives pour de futures recherches et applications.

A.2 APPROCHE DE MODELISATION

Notre objectif de modélisation consiste à concevoir un modèle générique de base de données sémantique et spatio-temporelle pour la représentation de mobilités en milieu urbain. La composante sémantique du modèle décline la notion de trajectoire telle qu'elle apparaît en domaine urbain, elle se complète de caractéristiques spécifiques permettant de qualifier également les dimensions spatiales et temporelles.

A.2.1 Modèle de mobilité

Nous proposons un schéma conceptuel de base de données des mobilités à partir de plusieurs schémas étroitement liés les uns aux autres (appelés modules ci-après). Les principes de conception de ce modèle sont inspirés d'une ontologie de trajectoire proposée par Yan et al. (2008), la construction du modèle est développée selon les principes de l'outil de conception MADS (*Modeling Application Data with Spatio-temporal features*) développé par Spaccapietra et Parent (2006). Ces trois schémas complémentaires sont les suivants:

- Un module **Trajectory** qui caractérise les déplacements urbains à partir du réseau de transport urbain.
- Un module **Transportation Network** qui représente les différents réseaux de transport de la ville.
- Un module **Travel Pattern** qui qualifie la sémantique des mobilités en milieu urbain.

Ces trois modules (Trajectory module, Transportation Network module, and Travel Pattern module) ne sont pas indépendants dans la mesure où les entités identifiées dans chacun des modules peuvent être en relation avec des entités d'autres modules. L'union de ces trois modules et schémas donne un modèle global de représentation **Travel Behavior schema** des mobilités dans la ville (Figure 1).

List. Une *Trajectory* est une séquence de *one-to-many TransportationSegments*. Une *Trajectory* est décrite par un attribut temporel, *Duration*, spécifié par un intervalle et dérivé respectivement par les instants caractérisant le premier noeud et le dernier noeud du premier *TransportationSegment* TR_1 et dernier segment *TransportationSegment* TR_n de cette trajectoire *Trajectory*.

Définition 1: Individual

Un *Individual* représente un humain qui se déplace le long d'un réseau de transport et réalise quelques activités dans la ville. Un *Individual* est modélisé par plusieurs attributs qui qualifient son identité et plusieurs de ses caractéristiques notamment son nom *Name* typé par un *String*, *Age* typé par un *Integer*, *Gender* typé par un *String* et deux valeurs possibles (i.e., *Male* or *Female*), *Occupation* typé par une *Enumeration* (i.e., liste de possibles emplois) et *FamilyRole* typé par une *Enumeration* (e.g., mère, père, enfant).

Définition 2 Trajectory

Une *Trajectory* se réalise à partir de plusieurs composants du réseau urbain et en particulier des types de transport utilisés. Une *Trajectory* est explicitement représentée par une séquence de points qualifiés temporellement (instant, latitude, longitude and altitude), et plus précisément modélisé par une séquence ordonnée de *TransportationSegments* : $Trajectory_i = [TS_1, TS_2, \dots, TS_n]$ où TS_1, TS_2, \dots, TS_n représente une séquence de *TransportationSegments*. Notons qu'un *TransportationSegment* peut être parcouru par plusieurs individus et par plusieurs trajectoires. Une *Trajectory* a implicitement un noeud de départ et un noeud d'arrivée respectivement donnés par le noeud de départ de son premier et dernier *TransportationSegments*.

A.2.3 Module Transportation Network

Le module *Transportation Network* représente le réseau urbain (réseau des rues, réseau de tramway, ...).

Définition 3: TransportationNode

Un *TransportationNode* modélise une intersection de ligne de transport ou un Point d'Intérêt (POI). Une intersection de ligne de transport matérialise une jonction entre plusieurs *TransportationSegments* (au moins deux). Un *POI* est un point d'intérêt matérialisé dans le réseau par une approximation de sa localisation.

Le type d'objet *TransportationNode* a un attribut spatial *Node* défini comme une géométrie de type *Point* qui représente sa localisation dans le réseau de transport. Un attribut Boolean *IsStop* distingue un *TransportationNode* qui modélise une intersection du réseau de transport (i.e., valué par 1) ou un *POI* (i.e., valué par la valeur nulle). *IsRoad*, *IsMetro* et *IsRailway* sont les trois attributs qui permettent de distinguer les réseaux de transport sous-jacent auquel un noeud appartient.

Définition 4: TransportationSegment

Un *TransportationSegment* modèle un lien entre deux *TransportationNodes* dans le réseau de transport. Un *TransportationSegment* est valué par un attribut spatial (i.e., de type segment de ligne) le long du réseau de transport, et terminé par deux *TransportationNodes*. La longueur de ce segment de ligne est aussi représentée par un attribut. L'attribut *Type* définit les différents types de *TransportationSegment*, c'est à dire route, metro ou train. Nous assumons que quand plusieurs *TransportationSegments* de différents *Types* se superposent ils sont représentés comme des objets différents. La relation entre un *TransportationSegment* et un *TransportationNode* est aussi définie comme une relation topologique *Touch* entre un segment de ligne et un point. Un *TransportationNode* est en relation avec one-to-many *TransportationSegment* alors qu'un *TransportationSegment* doit avoir deux *TransportationNodes* qui représentent respectivement l'origine spatiale et la destination de ce segment.

En sus des différents concepts et abstractions de modélisation présentés par la Figure 1, une contrainte spatio-temporelle est prise en compte afin de garantir la connectivité du réseau (d'autres contraintes dépendantes de l'application pourraient être définies selon les mêmes modalités):

Contrainte 1

Les successifs *TransportationSegments* qui composent une *Route* doivent être connectés, c'est à dire, soit le noeud de départ *TransportationNode* ou/et le noeud d'arrivée *TransportationNode* de ce *TransportationSegment* doivent être soit le noeud de départ ou de fin *TransportationNode* d'un autre *TransportationSegment* de cette *Route*.

A.2.4 Module Travel Pattern

Le module *TravelPatterns* permet de qualifier les primitives de déplacement extraites des trajectoires.

Définition 7: Route

Une *Route* représente un chemin suivi régulièrement par un ou plusieurs *Individuals*. Nous considérons un *Individual_i* en déplacement le long du réseau de transportation et qui donc génère une *Trajectory*. Une *Route* est généralisée à partir de *Trajectories* régulières. Une *Route* démarre à un *Node* et se termine à un *Node*. Les *TransportationSegments* et *TransportationNodes* qui constituent cette route sont implicitement ordonnés. Plus formellement une *Route* est modélisé comme une séquence de *TransportationSegments*. Afin de modéliser la spatialité d'une *Route*, nous introduisons le concept de *Path* défini comme une séquence de *TransportationSegments*.

$$Path_i = [TransportationSegment_1, TransportationSegment_2, \dots, TransportationSegment_n];$$

Afin de représenter des déplacements réguliers, nous introduisons la notion d'intervalle régulier qui représente des cycles d'intervalles ou en d'autres termes un intervalle temporel qui se répète :

$$TemporalCycle = \langle Regularity, Regular Interval \rangle$$

Un exemple de *TemporalCycle* : $TimeSegment = \langle workdays, 08h-09h \rangle$.

Une *Route* est donc définie par:

$$Route_i = \langle Path_i, TemporalCycle_i, TransportationMode \rangle$$

Alors qu'un *Individual* réalise une *Trajectory* donnée une seule fois, il peut réaliser un *Path* plusieurs fois, et une même *Route* régulièrement.

Définition 8: TravelPattern

Une *Route* d'un *Node N_l* vers un noeud *Node N_m* peut être partagée par plusieurs *Individuals* $\{i_1, \dots, i_n\}$. Une *Route* est le résultat implicite de plusieurs contraintes *Constraints* (*BehaviorActivity*, *TemporalConstraint*...). Au final, des *Routes* peuvent générer des patrons de déplacement *TravelPatterns* lorsqu'elles sont réalisées régulièrement ou même rarement (ce qui permettra d'étudier des tendances ou des particularités de déplacement).

Un même ensemble de *Trajectories* peut générer plusieurs *Routes* dont l'analyse va permettre d'étudier plusieurs modalités de patrons de déplacement *TravelPatterns*. Alors que les *Trajectories* et les *Routes* sont explicitement

représentés dans l'espace et dans le temps, les patrons de déplacement *TravelPatterns* sont dérivés de *Routes* à partir de contraintes de fréquence.

$TravelPattern = \langle Routine: [\langle Route_1 \rangle, \langle Route_2 \rangle, \dots, \langle Route_n \rangle], TemporalCycle, Frequency, TypeNumber \rangle$

Un patron de déplacement *TravelPattern* devrait de fait être la conséquence de plusieurs facteurs influençant des décisions et facteurs de déplacement d'un ou plusieurs *Individuals*. Par exemple, considérons l'exemple illustratif d'un jour de la semaine et différentes conditions météorologiques. Selon que le temps soit pluvieux ou pas un *Individual* choisira de se déplacer en voiture ou à pieds pour se rendre à son travail distant de quelques centaines de mètres. Ces contraintes et propriétés sont spécifiées par des attributs qui qualifient un patron de déplacement *TravelPattern* soit les attributs qui permettent d'identifier l'*Individual* et ces différentes contraintes. Ces contraintes incluent les contraintes environnementales (conditions météorologiques, paysages, état de la route ...), les contraintes temporelles (matin, après-midi, jours de la semaine ...), les contraintes spatiales (localisation, voisinage, points d'intérêts ...) et le but du déplacement (loisir, travail ...).

Généralement, un patron de déplacement *TravelPattern* sera généré par des contraintes internes et externes. Ce sont ces contraintes et ces influences qui vont constituer la trame des caractéristiques comportementales de notre approche de modélisation.

Les comportements de déplacements des *Individuals* traduisent souvent une large diversité d'actions et d'activités, reflets de différentes caractéristiques socio-culturelles et économiques. Bien que notre objectif consiste à développer une approche de modélisation permettant de prendre en compte les possibles propriétés influant des patrons de déplacement – un objectif très ambitieux – et afin d'illustrer le potentiel de notre approche, nous choisissons de retenir quelques propriétés élémentaires et attributs d'*Individual* souvent utilisés en sciences humaines comme le sexe, l'âge et l'emploi. Selon Hägerstraand (1970), les comportements humains de déplacement sont souvent contraints par des obligations sociales. Il est donc raisonnable de considérer lors de la modélisation d'un profil personnel de catégoriser socialement les *Individuals* afin de mieux générer des classes provenant de groupes homogènes et potentiellement exprimant des comportements convergents.

Par exemple, considérons les différentes caractéristiques économiques et occupations d'une famille. Les modalités de transport sont susceptibles de varier

selon les catégories sociales, les emplois occupés, les activités réalisées et la typologie du foyer. Dans la mesure où un même *Individual* est susceptible de réaliser plusieurs *Trajectories* sur une longue période de temps, ce même *Individual* est susceptible de générer plusieurs patrons de déplacement *TravelPatterns* sur cette même période.

Plusieurs facteurs influencent des décisions de déplacement d'un même *Individual* générant des patrons de déplacement *TravelPattern*. Nous classifions ces facteurs en quatre catégories qui prennent en compte des contraintes temporelles *TemporalConstraint*, des typologies d'activité *BehaviorActivity*, des contraintes environnementales *EnvironmentConstraint*, et des contraintes spatiales *SpatialConstraint* (d'autres catégories de contraintes pourraient être définies selon les mêmes modalités). Comme précisé plus haut, plusieurs caractéristiques sociales, telles par exemple que celles appréhendées par notre modèle, peuvent jouer un rôle dans la façon qu'aura un *Individual* de se comporter et se déplacer dans la cité. Dans l'instanciation de notre modèle et pratiquement un *Individual* est susceptible de « posséder » plusieurs *Trajectories* et ces *Trajectory* sont aussi susceptibles de générer des patrons de déplacements *TravelPatterns*. Ces catégories de contraintes matérialisent différents facteurs qui influencent ces comportements de déplacement. Dans le cas des contraintes spatiales *SpatialConstraint* nous considérons notamment des POInts d'Intérêt *POIs* et des quartiers *District*. Les *POIs* matérialisent des lieux qui présentent un intérêt pour un *Individual* ou des points de repère ou de convergence de ces déplacements (e.g., un supermarché, un bureau de tabac), ces *POIs* jouant le rôle de contraintes spatiales. Un autre exemple de contrainte spatiale provient de cette notion de *District* dont le rôle consiste notamment à caractériser les origines et les destinations des trajectoires de déplacement, cela potentiellement permettant d'étudier des phénomènes de déplacement origine-destination à différents niveaux d'agrégation dans la cité, les tendances générales et les modalités de transport utilisées, pour ne citer que quelques exemples de manipulation possible que ce type de modélisation permettra.

Considérons une deuxième catégorie de contrainte à savoir les contraintes temporelles *TemporalConstraints* comme des constituants de patrons de déplacement *TravelPattern*. Ces contraintes temporelles *TemporalConstraints* peuvent caractériser différentes périodes temporelles comme jour/nuit, heure de pointe, jours de semaine/jours de fin de semaine, journée de travail/journée de vacances etc. Ces types de contraintes temporelles jouent un rôle important dans

l'expression et la réalisation de patrons de déplacement. Ces contraintes temporelles *TemporalConstraints* peuvent être spécifiées selon les contraintes de l'application à partir de différents attributs et granules temporels.

Une activité comportementale *BehavioralActivity* est utilisée pour représenter le but du déplacement d'un *Individual*. Prenons l'exemple d'un déplacement sur son lieu de travail, dans ce cas la contrainte *BehavioralActivity* pourra être représentée par une valeur 'Travail' dans le module *TravelPattern*. Les activités d'un *Individual* peuvent être par exemple classées en plusieurs sous-catégories comme *PersonalActivity*, *FamilyActivity* and *BusinessActivity*. Typiquement ces différentes catégories d'activités représentent des classes de comportement bien particulières. Generally. Pour une activité *BusinessActivity*, minimiser le temps de déplacement est souvent un critère déterminant à appliquer ce qui ne sera pas forcément ou en tout cas pas avec la même magnitude dans le cas d'une *FamilyActivity*.

Le contexte environnemental est également une contrainte qui impacte les décisions et donc les patrons de déplacement *TravelPatterns*. Nous classifions en première approximation ces contraintes environnementales *EnvironmentalConstraints* en deux catégories: celles issues de considérations sociales *SocialEnvironment* et celles résultant de l'environnement naturel *NaturalEnvironment*. Les contraintes sociales *SocialEnvironment* modélisent des facteurs sociaux influent sur les décisions ou les conditions de déplacements alors que les contraintes environnementales prennent en compte elles des paramètres physiques de l'espace comme les conditions météorologiques, topographiques etc.. Toutes ces contraintes, à savoir *TemporalConstraints*, *EnvironmentalConstraints* et *BehavioralActivity* sont qualifiées par des attributs et des constructeurs de tuples permettant de les caractériser. A partir du modèle MADS, des liens de généralisation et de spécialisation is-a permettent de spécifier des *Supertypes* et des *Subtypes* qui peuvent respectivement généraliser ou spécialiser des attributs de leurs *Subtypes* et *Supertypes*, respectivement.

La dernière catégorie de contrainte spatiale *SpatialConstraint* modélise les propriétés spatiales qui peuvent avoir un impact sur les patrons de déplacement *TravelPattern* d'un *Individual*. Comme pour les catégories de contrainte précédemment définies, *SpatialConstraint* représente une catégorie de facteurs influençant des décisions de déplacement.

Une *SpatialConstraint* est susceptible d'être définie à partir de *POI* ou de *Districts* que les *Individuals* sont susceptibles de visiter ou de parcourir lors de leurs

déplacements. Les deux exemples de contraintes spatiales que nous avons retenus pour illustrer le potentiel de modélisation de notre approche sont les *POIs* et les *Districts*. Un *POI* est une localisation qui présente un potentiel intérêt pour un déplacement donné. Un *POI* est spatialement modélisé comme un point qui matérialise sa position dans l'espace. Ces localisations sont susceptibles de devenir la source ou la destination d'une *Trajectory*. Les attributs d'un *POI* peuvent permettre de rechercher les raisons qui ont conduit une catégorie de personnes à effectuer un tel déplacement d'un *POI A* à un *POI B*. Le rôle de ces districts est important dans le contexte de notre modèle dans la mesure où il va nous permettre de modéliser ces déplacements à différents niveaux d'abstraction, et le niveau des quartiers est l'un de ces niveaux présentant un intérêt pour l'étude des décisions qui ont générées tel ou tel déplacement. Plus pratiquement un *District* sera considéré comme une primitive spatiale permettant de structurer une ville donnée en plusieurs quartiers représentatifs. En conséquence un *District* et un autre type de contrainte qui joue un rôle important dans les typologies de comportement de déplacement.

Contrainte 2

L'union des Districts du contexte d'étude forme une partition de l'espace.

A.2.5 Relations entre Différents Modules

Comme dans toute approche de modélisation spatio-temporelle ou même plus générale, les relations Relationships sont des concepts de modélisation importants du schéma Travel Behavior schema. La plupart des connexions sémantiques entre les différents modules sont matérialisés par des relations qui relient des objets identifiés. Par exemple la relation Follow entre un objet Route et un objet Trajectory est une abstraction importante du module TravelPattern, tout comme la relation temporelle entre Route et TransportationSegment. Un patron de déplacement TravelPattern est caractérisé par différents facteurs dans le module TravelPattern. Comme pour la relation entre le module Travel Pattern Module et le module Trajectory, de multiples relations sont modélisées. La relation entre le Trajectory et le module de patron de déplacements Travel Pattern Module matérialise une relation indirecte entre TravelPattern, Individual et représente le fait que chaque Trajectory appartient à un Individual. L'abstraction Individual joue un rôle important dans la modélisation de comportements. Individual est une unité de modélisation essentielle de la modélisation de patrons de déplacement. Un Individual peut posséder une à plusieurs Trajectories qui peuvent refléter plusieurs types d'action et de déplacement.

Un District est en relation spatiale avec le réseau de transport. Considérant les possibles relations topologiques comme par exemple celles reliant les TransportationSegments aux Districts (Touch, Within and Intersect), il apparaît clairement que plusieurs configurations topologiques émergent et sont illustrées par le schéma suivant. Ces relations sont représentées dans la base de données initiale et par la relation topologique TopoGeneric (Figure 1). Les POIs jouent aussi un rôle comme les Trajectories. Un POI reflète une source, un intermédiaire ou un nœud de destination d'un déplacement. Un POI est connecté à son réseau et à partir d'une relation topologique avec un nœud du réseau, quand ce POI et les potentiels POIs sont disjoints. Au final, ces POIs jouent un rôle déterminant de qualité de service dans la matérialisation de routes optimales ou préférées.

A.3 ANALYSE DE PATRONS DE DEPLACEMENTS

A.3.1 Interrogations de comportements de déplacements (Travel Displacement Queries)

Les interrogations de comportements de mobilités (Travel behavior queries) devraient idéalement apporter des réponses ou des indications de réponses pour des interrogations orientées vers la recherche d'une meilleure compréhension des mobilités humaines. Il apparaît clairement qu'une catégorisation de ces interrogations devrait permettre de mieux les appréhender et les définir. Selon Dodge et al. (2008), les comportements de mobilité peuvent être catégorisés en trois classes: spatiales, temporelles et spatio-temporelles, auxquelles nous rajoutons la dimension sémantique. A partir de cette classification nous proposons d'étudier les types d'interrogations considérées comme des primitives de base qui pourront être d'une part spécifiées au sein d'un langage de manipulation de bases de données, et ensuite combinées pour exprimer des recherches à haute valeur ajoutée.

A.3.1.1 Niveau sémantique

A partir de ce niveau sémantique, les interrogations portent généralement sur des abstractions fondamentales exprimées à partir du « Qui », « Quoi », « Comment » ou « Pourquoi » et la manipulation de primitives de base. (Thériault and Claramunt, 1999)

A partir du module Travel Behavior, les objets sémantiques et leurs attributs sont pour l'essentiel déclinés dans le module Travel Pattern qui répertorie les possibles

facteurs qui influencent les décisions de déplacement et qui sont spécifiés à partir de trois catégories complémentaires : *Individual Profile*, *Behavioral Activity* and *Environmental Constraint*.

Bien que le module Travel Pattern Module décrive l'essentiel des données sémantiques, ou à minima les plus riches, d'autres opportunités d'extension à partir d'interrogations combinant les informations explicitement représentées au sein des autres modules. Par exemple, et à partir de l'objet *Individual*, il est possible de répondre à la question basique "*Qui est cet individu?*". Qualifier and reconnaître qui est un individu en déplacement devrait pouvoir permettre de répondre à de simples interrogations, rechercher des régularités de déplacement et explorer quelles seraient les raisons amenant à cette activité. Plus globalement, dériver une nouvelle part de connaissance des dynamiques et caractéristiques d'une mobilité, et en généralisant une telle approche à un ensemble d'individus se déplaçant dans la cité doit pouvoir permettre d'étudier les phénomènes de déplacement dans la cité. En rattachant par exemple de telles interrogations sémantiques à des détails de parcours et en particulier des *POIs* qui seraient parcourus ou certains types d'activités peut amener à caractériser un ensemble de patrons de déplacements, ou à mieux appréhender le caractère et/ou les particularités d'un ou d'un ensemble d'individus.

A.3.1.2 Niveau temporel

Les interrogations appliquées au niveau temporel doivent privilégier l'étude de régularités ou d'irrégularités des mobilités d'un ou de plusieurs individus, des origines et/ou des destinations de déplacement, les temps passés à une destination donnée pour réaliser une certaine activité, les temps de parcours pour ne citer que quelques exemples. Ces interrogations sont particulièrement importantes pour des études de transport et de planification en milieu urbain. Du point de vue des abstractions de base utilisées ces interrogations utilisent le plus souvent les qualités « Quand », « Combien de temps », « Combien de fois » et peuvent s'appliquer préférentiellement à partir du module Module.

A.3.1.3 Niveau spatial

Les interrogations portant sur cette dimension spatiale manipulent des données de localisations et des géométries. En dehors d'expressions basiques telles que « Où se situe cet individu », d'autres catégories d'interrogations peuvent être appliquées à partir de métriques, et d'opérateurs topologiques et de direction, ou leur combinaisons.

A.3.1.4 Niveau spatio-temporel

Les interrogations combinant les niveaux spatiaux et temporels combinent ces dimensions à partir notamment d'objets représentés dans le module Travel Behavior. Ces interrogations manipulent les abstractions « Où et Quand » et peuvent typiquement porter sur des propriétés de trajectoires (e.g., origine et destination, localisation, parcours) ou sur des contraintes spatiales SpatialConstraints et temporelles TemporalConstraints introduites dans les sections précédentes. Un exemple de patron qui peut être analysé aborde des propriétés fonctionnelles. Par exemple, une comparaison explorerait les caractéristiques spatiales des déplacements journaliers des Trajectory d'un Individual et les Districts croisés lors de ces déplacements. D'autres exemples plus significatifs pourraient par exemple rechercher les différences et les similarités de déplacement entre les trajectoires d'un même individu à plusieurs temps donnés, ou entre plusieurs individus, ou des fréquences de déplacements dans certaines conditions temporelles. Ces mêmes tendances de déplacement pourraient même être étudiées à différents niveaux de granularité temporelle et échelle spatiale.

A.3.1.5 Niveaux sémantiques, spatiaux et temporels

Afin de mieux comprendre et appréhender les comportements de mobilité dans la cité il est naturel de vouloir combiner les dimensions sémantiques, spatiales et temporelles. Les opportunités d'exploration sont alors démultipliées. Elles peuvent aborder et étudier les conditions de déplacement d'un ou plusieurs *Individual*. Les interrogations portant sur le niveau d'un *Individual* sont susceptibles de rechercher des régularités ou des irrégularités. Plus généralement les interrogations portant sur plusieurs *Individuals* pourraient par exemple recherche « qui sont les individus qui sont passés par telle route à un temps donné » ou « quels sont les lieux qu'un individu donné a visité à tel moment »

A.4 IMPLEMENTATION ET VALIDATION

La mise en œuvre de ce modèle de comportement de voyage est réalisée à partir d'un système de base de données spatiales PostgreSQL/PostGIS. PostgreSQL est une base de données Open source qui a l'avantage de permettre d'utiliser des extensions spécifiées par l'utilisateur. PostGIS est l'extension spatiale de la base de données

relationnelle objet PostgreSQL qui fournit des types de données spatiaux, des index spatiaux et des fonctions de manipulation de données spatiales. Nous avons donc développé une base de données de représentation de mobilités à partir de cet environnement PostgreSQL/PostGIS et instanciées à partir du modèle de données mentionné dans les sections précédentes. Sont notamment spécifiées les objets trajectoires, les individus et les relations les associant. Comme dans de nombreuses approches de modélisation de base de données, toutes les entités identifiées par le modèle de mobilité sont « mappées » vers des tables relationnelles. Des index spatiaux GIST-index et temporels b-Tree offerts par PostgreSQL/PostGIS ont été appliqués pour accélérer les requêtes spatiales et temporelles. Des opérations spécifiques de manipulation de données ainsi que les opérations temporelles spatiales ont été implémentées en tant que fonctions au niveau de l'interface.

A.4.1 Conception de Fonctions

Les fonctions constituent des éléments de base des requêtes relativement évoluées des langages d'interrogation de données. Une base de données spatiales doit habituellement offrir une série d'opérateurs et de fonctions qui manipulent des données quantitatives (e.g., métrique) ou qualitatives (e.g., topologique). De nombreux opérateurs, fonctions spatiales et temporelles de base sont le plus souvent disponibles dans les systèmes de base de données spatiales actuels, tel qu'Oracle Spatial, PostGIS etc. Ces fonctions comprennent notamment les métriques et les constructeurs de géométrie (e.g., distance, longueur, buffer), les opérateurs topologiques (e.g., *cross*, *within*, *Touch*). Cependant, ces opérateurs et ces fonctions ne sont pas toujours suffisants pour des applications spécifiques à sémantique de haut niveau, comme pour l'analyse à effectuer par notre étude orientée vers l'analyse des mobilités. Il en résulte que plusieurs catégories de fonctions spécifiques doivent être dérivées et agrégées et définies à partir de sémantiques explicites pour les utilisateurs et qui s'adapteront aux fonctions requises par notre approche. De telles fonctions devraient directement et explicitement faciliter la recherche de modalités de mobilités de certaines personnes, et devraient même être construites pour à la fois traiter et manipuler des données spatiales, temporelles et sémantiques.

Avec l'objectif de s'appliquer préférentiellement à partir de notre modèle conceptuel et logique de représentation des mobilités, les fonctions à construire ont pour objectif d'aider à découvrir des comportements de mobilité dans la ville. La conception de ces fonctions est basée sur les trois catégories décrites dans la section précédente. Afin de tenir compte de la complexité des analyses possibles de

comportements de mobilité, la plupart des fonctions identifiées combinent les dimensions spatiales, temporelles et sémantiques. Des opérateurs géométriques et topologiques spécifiques ont été dérivés d'une catégorisation suggérée dans une précédente recherche (Wu et al., 2014). Le tableau 1 donne une liste des fonctions conçues à ce stade, y compris leur signature et une brève description.

Table A. 1 Liste de Fonctions

No.	Name	Signature	Type	Description
1	Ind_At_Instant	Individual x Timestamp-> Geometry(POINT)	Manipulation spatio-temporelle	Retourne la position d'un individu à un temps donné
2	Ind_Traj_During_Period	Individual x Time_Interval-> geometry(LINESTRING)	Manipulation spatio-temporelle	Retourne les segments de transport qu'un individu réalise pour un intervalle temporel donné
3	Ind_Enter_At	Individual x Geometry-> Timestamp	Interrogation topologique	Retourne le temps quand un individu rentre dans une Geometry(LINESTRING or POLYGON)
4	Ind_Exit_At	Individual x Geometry-> Timestamp	Interrogation topologique	Retourne le temps quand un individu est localisé dans une Geometry(LINESTRING or POLYGON)
5	Ind_Confine_In	Geometry(Polygon) x Time_Interval -> Set(Trajectory,Individual)	Interrogation topologique	Retourne les trajectoires d'un individu confiées dans un MBB spatio-temporel
6	Traj_MBR	set(Trajectory) -> Region	Manipulation spatiale	Retourne le minimum rectangle englobant (MBR) d'un ensemble de trajectoires
7	Traj_Pass_District	Trajectory x District -> Boolean	Interrogation topologique	Détermine si une trajectoire <i>CrossIn</i> et puis <i>CrossOut</i> un district

8	Traj_Pass_Transport	Trajectory x (TransportationSegment) -> Boolean	Interrogation topologique	Détermine si une trajectoire <i>CrossIn</i> et puis <i>CrossOut</i> un segment de transport
9	Traj_Nearest_POIs	StartTraj EndTraj(Trajectory) x Distance -> set(POI)	Buffer	Trouve les points d'intérêt POIs à une certaine distance du début ou la fin d'une trajectoire
10	Traj_Direction	Trajectory -> Bearing	Interrogation spatiale	Retourne la direction d'une trajectoire
11	Traj_Co-direction	Trajectory x Trajectory - > Boolean	Interrogation spatiale	Determine si deux trajectoires sont orientées dans la même direction
12	Ind_Find_Route	Individual x Time_Interval-> Set(Route)	Aggrégation	Retourne les routes extraites des trajectoires d'un individu pour un intervalle de temps donné
13	Traj_Fit_Route	Trajectory x Route -> Boolean	Analyse spatio- temporelle	Détermine si une trajectoire suit une route donnée
14	Ind_Share_Path	set(individual) -> set(path)	Aggrégation	Retourne les chemins partagés par plusieurs trajectoires
15	Trans_Avg_Speed	set(TransportationSegme nts) x Timestamp -> Speed	Analyse spatio- temporelle	Retourn la vitesse poyenne des trajectoires passant par un segment de transport et pour un intervalle de temps donné

A.4.2 Données expérimentales de trajectoires

Les requêtes implémentées sont exécutées à partir d'une large base de données de trajectoires extraites de la ville de Pékin. Ce jeu de données de trajectoires dérivées de signaux GPS a été collecté par le projet Geolife (Microsoft Research Asia) (Zheng et al., 2008; Zheng et al., 2009; Zheng et al., 2010) par 178 utilisateurs sur

une période de quatre ans d'avril 2007 à octobre 2011. Une trajectoire GPS dans ce référentiel est représentée par une séquence de points estampillés dans le temps et dont les attributs spatiaux sont la latitude, la longitude et l'altitude. Ce jeu de données contient 17 621 trajectoires couvrant une distance totale de 1 251 654 kilomètres et une durée totale de 48 203 heures. Ces trajectoires ont été enregistrées par différents utilisateurs et capteurs GPS ou de téléphones équipés de GPS, et couvrent une grande variété de taux d'échantillonnage : 91% des trajectoires sont enregistrées dans une représentation dense, par exemple toutes les 1 ~ 5 secondes ou tous les 5 ~ 10 mètres par point.

Ce jeu de données Geolife a enregistré un large éventail de mouvements dans la ville de Beijing, en qualifiant notamment les modes de transport utilisés. Alors que certains utilisateurs ont porté un enregistreur GPS pendant quelques années, d'autres n'ont utilisé cette configuration similaire que pour quelques semaines. Les noms des individus dans la base de données et les requêtes sont simulés dans la mesure où les informations personnelles des utilisateurs ne sont pas disponibles. Les données de transport et d'autres données spatiales de la ville de Pékin sont issues de données spatiales du domaine public, y compris les réseaux routiers et de métro, les districts, et une série de POI extraits du service de cartographie Baidu (Lbsyun Baidu, 2016).

A.4.3 Extraction des trajectoires

Afin d'associer les trajectoires extraites du projet Géolife au réseau de transport, un processus algorithmique de mise correspondance de ces trajectoires avec le référentiel cartographique a été mis en œuvre. L'implémentation de cet algorithme est principalement basée sur une méthode basée sur des contraintes proposée par Brakatsoulas et al. (2005). Le principe qui sous-tend cette approche consiste à faire correspondre les segments de trajectoires vers les segments du réseau en tenant compte de la connectivité des graphes et des contraintes spatio-temporelles (i.e., la distance couverte par une trajectoire entre deux segments de transport doit être plausible). Ce processus a été exécuté par des fonctions pl/pgsql spécifiquement développées à partir de PostgreSQL. Cette approche permet de générer les segments de transport *TransportationSegments* et les nœuds de transport *TransportationNodes* modélisant le réseau de transport.

A.4.4 Résultats

Nous introduisons une série de requêtes illustrant le potentiel de notre approche de modélisation. Ces requêtes combinent les fonctionnalités du langage

PostgreSQL/PostGIS avec les fonctions définies dans les sections précédentes. Ces requêtes manipulent les dimensions sémantiques, spatiales et temporelles.

Requête 1: Quels sont les segments de route que Li Lei a effectués de 7h à 10h ? Et quels sont les segments de route que Li Lei a régulièrement répété de 18h à 21 heures ?

```
Select Ind_find_route(id, '7:00:00', '10:00:00') from IndividualTable where name = 'Li Lei' ;
```

```
Select Ind_find_route(id, '18:00:00', '21:00:00') from IndividualTable where name = 'Li Lei';
```

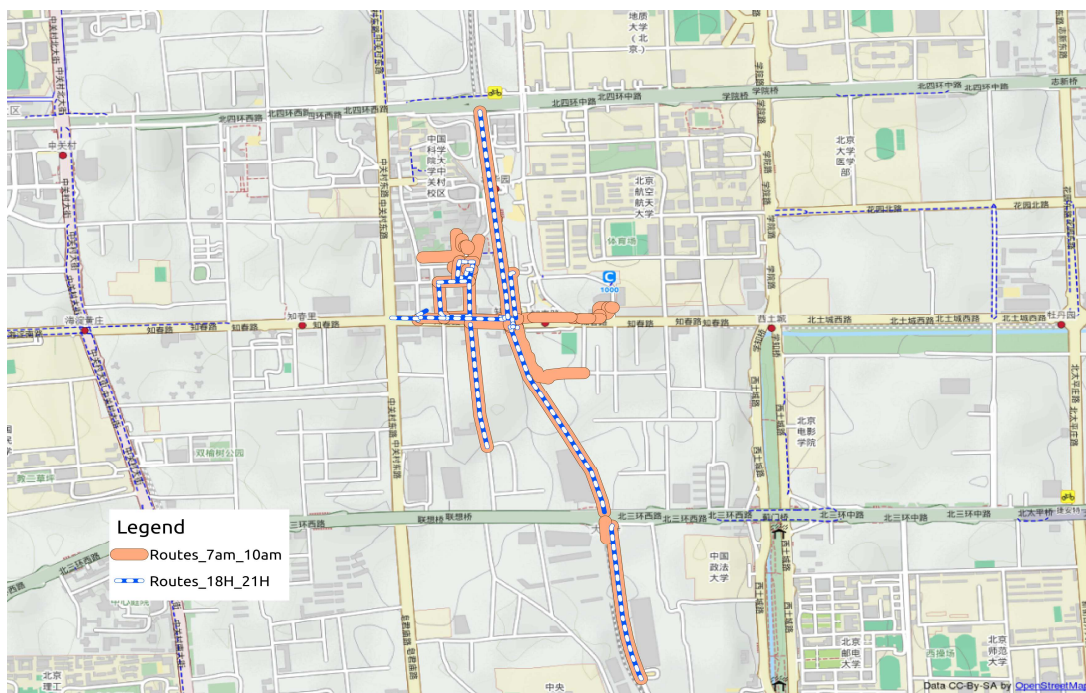


Figure A. 2 Segments de transportation régulièrement répétés par Li Lei le matin et le soir

La fonction *Ind_Find_Route* (*IndividualId*, *EndTime*) est encapsulée dans ces deux requêtes dont les résultats sont illustrés par la Figure 2. Cette fonction retourne les itinéraires suivis par Li Lei pendant les intervalles de temps donnés. Ces deux requêtes font implicitement une différence entre les itinéraires effectués par une personne tôt le matin et tard dans l'après-midi. Il apparaît à partir des résultats de la requête illustrés par la figure 2 que les itinéraires effectués en fin d'après-midi sont plus flexibles que ceux effectués tôt le matin. Cela pourrait refléter un ensemble

relativement important d'activités effectuées par cette personne après son travail, et d'autre part beaucoup moins de flexibilités pour celles effectués le matin et de plus réguliers patrons de trajectoires. Ce type de requête illustre comment l'exploration de patrons de trajectoire peut aider à l'exploration de comportements type..

Requête 2: Quels sont les segments que Li Lei répète régulièrement pendant les jours de la semaine et les week-ends ?

```
Select Ind_find_route(id, 'weekday') from IndividualTable where name = 'Li Lei' ;
```

```
Select Ind_find_route(id, 'weekend') from IndividualTable where name = 'Li Lei';
```

La fonction *Ind_find_route* est appliquée dans ces deux requêtes. Elles recherchent les itinéraires effectués par cette personne en semaine et le week-end. Elles permettent de faire une différence entre les comportements de déplacement en semaine et les week-ends et d'en analyser les différences. La figure 3 illustre la palette des itinéraires suivis. Il apparaît clairement que la plage géographique des itinéraires suivis pendant les jours de semaine est plus petite que ceux suivis pendant les week-ends. Cela traduit des patrons de mobilité et des régularités d'activité en semaine et à contrario une plus étroite gamme d'activité pendant les week-ends mais avec une plus grande étendue géographique.

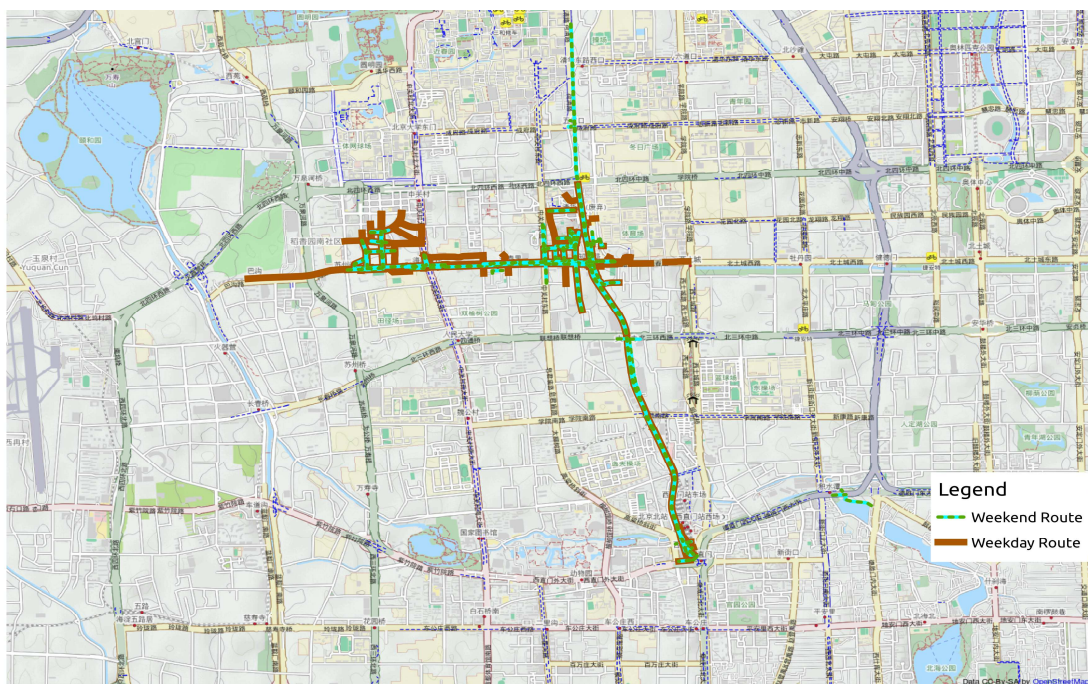


Figure A. 3 Routes les week-ends et en semaine pour un individu donné

Requête 3: Quelle est l'étendue spatiale des mobilités effectuées par Wang Hua et Li Lei à partir de 9h le 21 septembre 2008?

```
Select      Traj_MBR(union(trajjectory.id))from      TrajectoryTable      traj,
IndividualTable indiv where traj.individualid = indiv.id and (indv.name = 'Wang
Hua' or indiv.name = 'Li Lei') and extract(hour from starttime)>9;
```

La fonction Traj_MBR (TrajectoryId) est encapsulée dans cette requête. Cette fonction renvoie le *Minimum Bounding Rectangle* (MBR) des trajectoires données. L'objectif de cette requête est de révéler l'étendue géographique respective des trajectoires effectuées par certaines personnes pour un intervalle de temps donné. Plutôt que de comparer la géométrie exacte des trajectoires effectuées par une personne, l'idée sous-jacente à cette fonction consiste à comparer visuellement les régions couvertes, ces régions étant dérivées par un MBR. Lorsque certains individus partagent certaines régions géographiques, cela pourrait révéler des modèles de comportement remarquables et des intérêts communs lorsqu'ils se déplacent dans la ville et donc des modalités de co-voiturage ou de transport collectif à étudier. La figure 4 montre le résultat de cette requête appliquée à deux individus donnés. Dans ce cas précis, les figures qui apparaissent montrent que ces deux individus partagent une localisation proche à la fin de leurs trajectoires le matin. Il serait alors possible que les deux individus soient des collègues de travail potentiels.

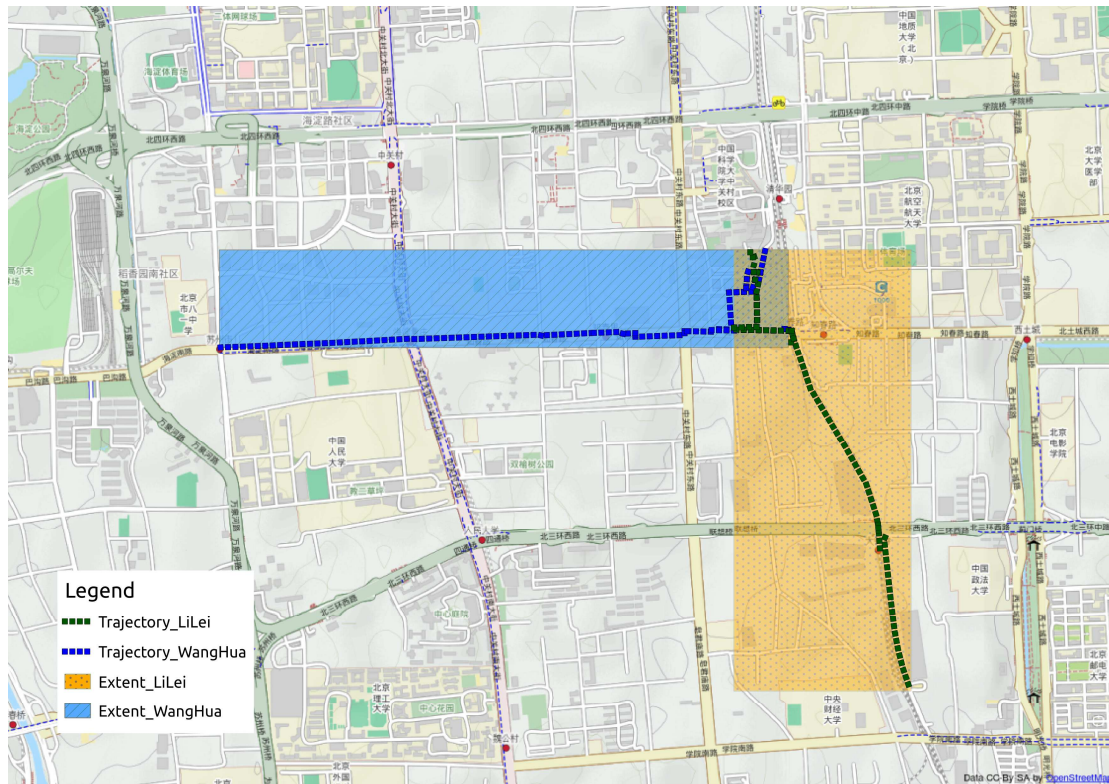


Figure A. 4 Couverture spatiale des trajectoires effectuées par des individus donnés pour un temps donné

Requête 4: Recherche les districts de provenance des personnes qui prennent la ligne de métro 10 entre 6 heures et 9 heures

```
Select district.id, district.geom from district, TrajectoryTable trj,
TransportationSegmentTable seg, TransportationTable trans where
Traj_PassTransport (trj.id, seg.id) and extract(hour from trj.starttime) between 6am
and 9am and seg.roadid = trans.id and trans.name = 'subway line 10' and
st_within(trj.startpt,district.geom);
```

La fonction *Traj_PassTransport (Trajectory.Id, TransportationSegment.ID)* est appliquée à cette requête. Cette fonction booléenne retourne 'vrai' lorsqu'une trajectoire donnée passe par un segment de transport *TransportationSegment* donné. L'objectif de cette requête est d'explorer et de rechercher les lieux d'où les personnes prennent la ligne de métro 10 le matin. Dans la ville de Beijing, une telle requête pourrait permettre de rechercher des configurations de patrons de transport, une information potentiellement utile pour la planification en transport. Le résultat de

cette requête est illustré à la figure 5. Il apparaît que les individus qui prennent la ligne de métro 10 le matin proviennent principalement du nord-ouest de la ville.

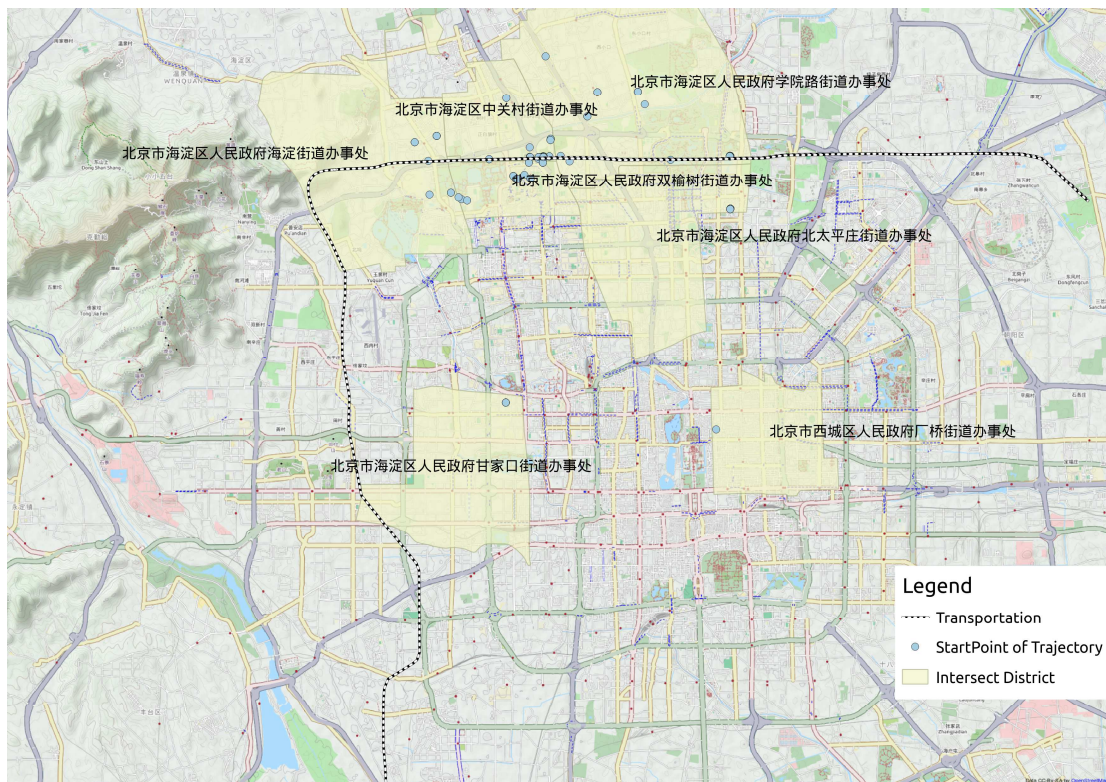


Figure A. 5 Districts matérialisant des débuts de trajectoires de la ligne de métro 10

Requête 5: Recherche les itinéraires les plus réguliers suivis par les personnes dans l'après-midi de 5 heures à 7 heures. Et qui sont globalement les chemins les

plus populaires suivis par la plupart des individus dans l'après-midi de 5 heures à 7 heures ?

```
Select rou.id, num, trans.name, trans.type, seg.geom from (select sum(array_length(trajs,1)) num, rou.segmentid from RouteTable rou group by segmentid order by num desc limit 15), TransportationTable trans, TransportationSegmentTable seg where rou.segmentid = seg.id and trans.id = seg.roadid;
```

```
Select rou.id, num_indiv, trans.name, trans.type, seg.geom from (select count(rou.individualid) indiv, rou.segmentid from RouteTable rou group by segmentid order by indiv desc limit 15), TransportationTable trans, TransportationSegmentTable seg where rou.segmentid = seg.id and trans.id = seg.roadid;
```

Ces deux requêtes appliquent des agrégations spatio-temporelles. Les résultats sont illustrés par les Figures 6 et 7. Le but de ces requêtes est de trouver les itinéraires les plus populaires dans la ville pour un intervalle de temps donné ainsi que les plus densément réalisés. Ces requêtes devraient révéler à la fois des parcours fréquents et ceux qui sont très fréquentés à certains temps. Les résultats de ces requêtes révèlent les itinéraires les plus fréquemment exécutés par des individus (première requête, i.e., les 15 premiers itinéraires) et les plus occupés (deuxième requête). La Figure 6 montre les itinéraires les plus densément parcourus, tandis que la Figure 7 décrit les parcours les plus fréquents et révèle des régularités d'itinéraires. Il semble que les itinéraires les plus populaires soient relativement dispersés dans la ville, alors que les itinéraires fréquentés par la plupart des individus sont plus concentrés dans une même région. Le nombre d'individus agrégés dans la Figure 7 est relativement faible dans la mesure où peu de personnes suivent des itinéraires réguliers. Cela reflète la diversité des comportements de mobilité dans cas.

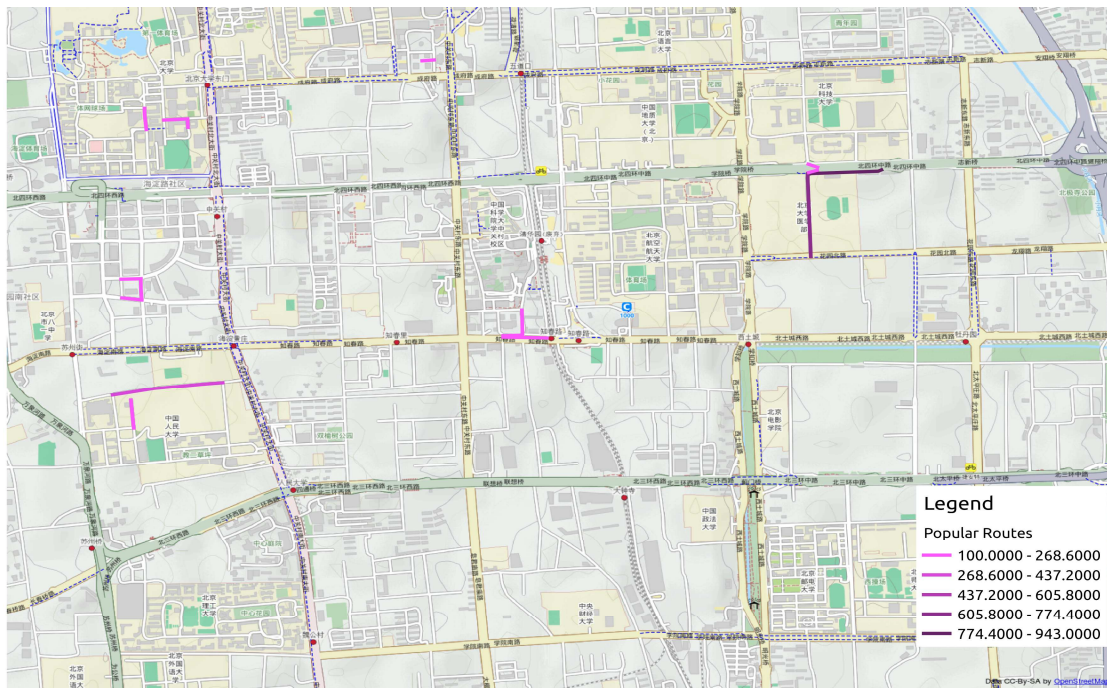


Figure A. 6 Routes densément fréquentées

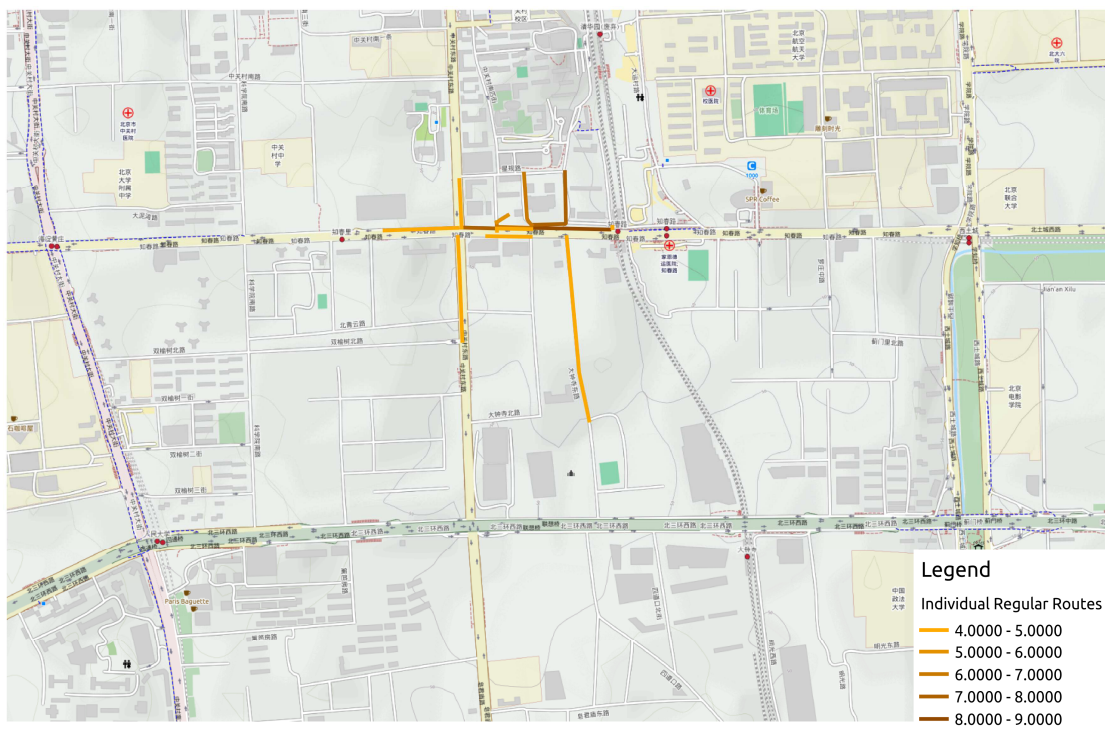


Figure A. 7 Routes populaires

Requête 6: Où la personne nommée «Li Ming» commence habituellement et termine ses déplacements avant 9:00 le matin? Quelles sont les destinations possibles ?

```
Select Traj_Nearest_POIs((select endpt from TrajectoryTable traj,
IndividualTable indiv where indiv.id = traj.individualid and indiv.name = 'Li Ming'
and extract(hour from traj.starttime) < 9),50);
```

La fonction *Traj_Nearest_POIs* (*Trajectory.StartPT/EndPT*, *Radium*) est appliquée à cette requête. Cette fonction renvoie le point d'intérêt le plus proche du début ou de la fin de la trajectoire pour une distance donnée. Les résultats de la requête sont illustrés par la Figure 8. Ils montrent que les destinations matinales sont relativement limitées à une petite région. On pourrait conclure que ce modèle régulier dans le voisinage de son domicile est susceptible d'être une trajectoire vers son lieu de travail. Les *POIs* proches de cette zone de destination pourraient représenter les lieux de travail possibles de cet individu. On pourrait même explorer les lieux de travail possibles selon les catégories de *POIs* présentes dans cette région. D'autre part, les trajectoires pratiquées le matin pour atteindre le lieu de travail potentiel reflètent une relative diversité, une raison possible étant des arrêts probables pour par exemple prendre son petit déjeuner.

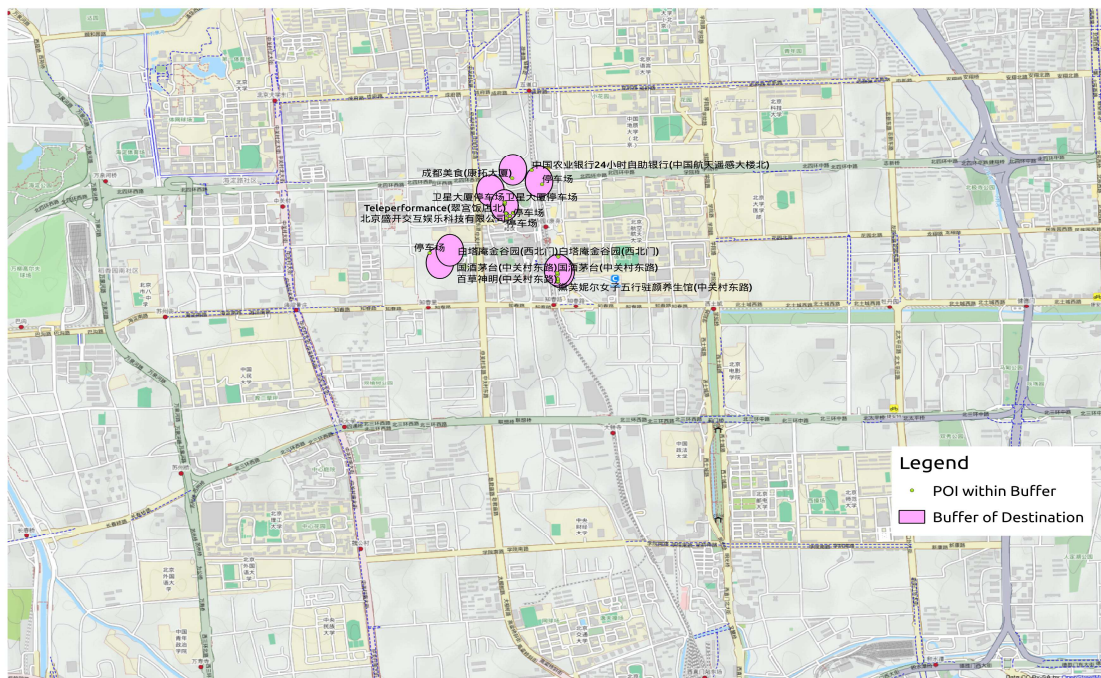


Figure A. 8 Points d'intérêts POIs proches de destinations en matinée

Au vu des résultats de l'ensemble de ces requêtes précédentes qui ont un caractère à la fois illustratif et exploratoire, il apparaît que dans de nombreux cas, l'analyse des modèles de trajectoire peut révéler des habitudes de mobilités voire certains comportements sociaux, et ce, même si la sémantique des données de base telles que fournies par la base de données Geolife est relativement limitée.

A.4.5 Evaluation des performances

Les performances de calcul des requêtes présentées ci-dessus montrent que cette implémentation basée sur notre modèle de comportement de mobilités est capable de manipuler efficacement de grandes données de déplacement extraites des données GPS brutes. Toutes les requêtes ont été exécutées sur un processeur Intel Core i7 avec 8 Go de RAM. Les délais d'exécution varient de quelques millisecondes à quelques centaines de millisecondes, ce qui est largement acceptable au niveau d'une interface et d'un utilisateur. Par exemple, les temps d'exécution des deux requêtes 5 sont les plus rapides (i.e., temps d'exécution de 325 ms et 3.149 ms). Ces temps d'exécution très rapides sont le résultat de la manipulation des seuls attributs sémantiques. La requête 3 (i.e., temps d'exécution de 426,244 ms) et la requête 6 (i.e., temps d'exécution de 608.800 ms) sont légèrement plus lentes, bien qu'elles restent encore rapides. Cela s'explique par l'application d'une jointure de table encapsulée dans les fonctions prédéfinies *Traj_MBR* et *Traj_Nearest_POIs*. Inversement, la requête 1 (i.e., temps d'exécution de 1331,423 ms et de 1094,187 ms), la requête 2 (i.e., temps d'exécution de 12707,713 ms et de 6258,714 ms) et la requête 4 (i.e., temps d'exécution 11704.87 ms) sont les requêtes les plus lentes car elles impliquent des fonctions qui encapsulent des opérateurs spatiaux sur plusieurs tables. Ces temps d'exécution restent globalement acceptables bien que des opérateurs spatiaux, sémantiques et temporels soient appliqués, et dont les résultats sont affichés visuellement au niveau de l'interface. Au final, il apparaît clairement que l'extensibilité de PostgreSQL/PostGIS s'est avérée favorable à la mise en œuvre de ce modèle de comportement de mobilités ainsi que les fonctions spécialisées identifiées.

A.5 CONCLUSION

La recherche développée dans cet article introduit une approche conceptuelle de base de données spatio-temporelles dont l'objectif est de modéliser les trajectoires humaines dans un environnement urbain et de fournir des capacités de manipulation et d'analyse de ces données. Le modèle de données de mobilités propose plusieurs concepts de modélisation à des niveaux sémantiques, spatiaux et temporels. Le modèle identifie plusieurs notions primitives telles que les abstractions d'individus, de trajectoires et de routes, l'ensemble étant relié par une série de relations et une association aux réseaux urbains sous-jacents. Les principales caractéristiques des réseaux urbains et de transport sont prises en compte, tout comme une série de contraintes sémantiques, temporelles et spatiales. A partir d'une implémentation réalisée sur les systèmes de base de données PostGIS et PostgreSQL, plusieurs fonctions spécifiques ont été identifiées et mises en œuvre. Ces fonctions permettent de manipuler les dimensions sémantiques, temporelles et spatiales à partir de modalités proches des besoins utilisateurs, et ont pu être expérimentées et mises en œuvre. Les principes de cette approche peuvent être étendus à d'autres opérateurs de manipulation de données génériques et spécifiques, susceptibles d'être distribués et partagés avec d'autres utilisateurs.

A gray square graphic containing the word "Appendix" in a white serif font at the top and a large white letter "B" in the center.

List of Publications

Parts of this work presented in this thesis are based on papers published or accepted for publication in:

M, Jin and C, Claramunt. "Discovering Human Travel Behaviour in Urban Region". Proceedings of the 20th AGILE Conference, Workshop of Bridging space, time and semantics in GIScience, 2017. http://geosem.ntua.gr/images/Discovering_Human_Travel_Behaviour_in-Urban_Region.pdf

Chapter III: Section 2,3

Chapter IV: Section 2, 3

M. Jin, C. Claramunt and T. Wang, "A Map-Matching Approach for Travel Behavior Analysis," 2017 4rd International Conference on Systems and Informatics (ICSAI), Hangzhou, 2017 (Accepted)

Chapter III: Section 2

Chapter IV: Section 3

Chapter V: Section 4, 5



Short Curriculum Vitae

Name: Meihan JIN

Place and Date of Birth: Jilin(China), 16/06/1988

EDUCATION:

- **2015-2017: Naval Academy Research Institute**

Ph.D candidate in Geomatics

Thisis: A Conceptual and Semantic Modelling Approach for the Representation and Exploration of Human Trajectories

- **2014-2015: Central South University, China**

Ph.D candidate in Cartography and Geographical Information Engineering

- **2011-2014: Central South University, China**

M.S in Geographic Information System

Thesis: Native Trajectory Database Based on PostgreSQL/PostGIS

GPA: 3.60

- **2007-2011: Central South University, China**

B.S in Geographic Information Science

Thesis: Point Cloud Spatial Database Extension.

GPA: 3.37

WORKING &PROJECT EXPERIENCE:

- Participated ESRI Programming Competition in China, Jun. 2009--Oct. 2009
Our team of three won the third prize with an online farm game based on GIS. (.Net, Openlayers, ArcServer, SQLserver)
- Intern at SoWell Computer Technic Company, Jul. 2011--Sep. 2011
Helped to design software for the City Planning Bureau. (Program of City Planning Project Management using .Net and AutoCAD)
- Science Innovation Program Funded by the university for my research program, Mar. 2013--Mar. 2014

Working on the project on algorithm for traffic congestion anticipation.
(PostgreSQL/PostGIS extension)

PUBLICATIONS :

[1]M, Jin and C, Claramunt. "Discovering Human Travel Behaviour in Urban Region". Proceedings of the 20th AGILE Conference, Workshop of Bridging space, time and semantics in GIScience, 2017.

http://geosem.ntua.gr/images/Discovering_Human_Travel_Behaviour_in-Urban_Region.pdf

[2]M. Jin, C. Claramunt and T. Wang, "A Map-Matching Approach for Travel Behavior Analysis," 2017 4rd International Conference on Systems and Informatics (ICSAI), Hangzhou, 2017 (Accepted)

REFERENCES

- Allen, E., Edwards, G., & Bédard, Y.(1995) . Qualitative causal modeling in temporal GIS. *Spatial Information Theory A Theoretical Basis for GIS*. Springer Berlin Heidelberg. 397-412.
- Alvares, L. O., Bogorny, V., de Macedo, J. A. F., Moelans, B., & Spaccapietra, S. (2007, November). Dynamic modeling of trajectory patterns using data mining and reverse engineering. In *Tutorials, posters, panels and industrial contributions at the 26th international conference on Conceptual modeling-Volume 83* (pp. 149-154). Australian Computer Society, Inc..
- Alvares, L. O., Bogorny, V., Kuijpers, B.(2007) A Model for Enriching Trajectories with Semantic Geographical Information. *Proceedings of the 15th International Symposium on Advances in Geographic Information Systems, ACM GIS 2007*
- Armstrong, M. P. (1988). *Temporality in spatial databases*.
- Andrienko, N., Andrienko, G., Pelekis, N., & Spaccapietra, S. (2008). Basic concepts of movement data. In *Mobility, data mining and privacy* (pp. 15-38). Springer Berlin Heidelberg.
- Axhausen, K. W., Zimmermann, A., Schönfelder, S., Rindsfuser, G., & Haupt, T. (2002). Observing the rhythms of daily life: A six-week travel diary. *Transportation*, 29(2), 95-124.
- Bishr, Y. (1998). Overcoming the semantic and other barriers to GIS interoperability. *Int. J. Geographical Information Science*, 12(4), pp. 299-314.
- Brakatsoulas, S., Pfoser, D., Tryfona, N.(2004) Modeling, Storing and Mining Moving Object Databases. *Database Engineering and Applications Symposium, 2004. IDEAS '04. Proceedings. International:pp 68-77*.
- Brakatsoulas, S., Pfoser, D., Salas, R., & Wenk, C. (2005, August). On map-matching vehicle tracking data. In *Proceedings of the 31st international conference on Very large data bases* (pp. 853-864). VLDB Endowment.
- Bogorny, V., Kuijpers, B., and Alvares, L. O. (2009) ST-DMQL: A Semantic

Trajectory Data Mining Query Language. *International Journal of Geographical Information Science* vol. 23.

Buliung, R. N., Roorda, M. J., & Remmel, T. K. (2008). Exploring spatial variety in patterns of activity-travel behaviour: initial results from the Toronto Travel-Activity Panel Survey (TTAPS). *Transportation*, 35(6), 697-722.

Buchin, K., Buchin, M., Gudmundsson, J., Löffler, M., & Luo, J. (2008, December). Detecting commuting patterns by clustering subtrajectories. In *International Symposium on Algorithms and Computation* (pp. 644-655). Springer Berlin Heidelberg.

Buchin, K., Buchin, M., Gudmundsson, J., Löffler, M., & Luo, J. (2011). Detecting commuting patterns by clustering subtrajectories. *International Journal of Computational Geometry & Applications*, 21(03), 253-282.

Camossi, E., Bertolotto, M., & Bertino, E. (2006). A multigranular object - oriented framework supporting spatio - temporal granularity conversions. *International Journal of Geographical Information Science*, 20(05), 511-534.

Chen, C. X. (2008). Spatio-temporal databases. In *Encyclopedia of GIS* (pp. 1121-1121). Springer US.

Claramunt, C., and Thériault, M. (1995). "Managing time in GIS an event-oriented approach." *Recent Advances in Temporal Databases*. Springer London. 23-42.

Claramunt, C. and Mainguenaud, M., 1996, A spatial data model for navigation knowledge, *Advances in GIS II*, Kraak, M. J. and Molenaar, M. eds, Taylor and Francis, Delft, the Netherlands, pp. 767-784

Claramunt, C. and Mainguenaud, M., 1999, A revisited projection operator for the manipulation of networks within GIS, *Informatica, special issue on spatial data management*, F. E. Petry, M. A. Cobb and K. B. Shaw (eds.), 23(2), pp. 187-202

Claramunt, C., and Theriault, M. (1996). "Toward semantics for modelling spatio-temporal processes within GIS." *Advances in GIS Research I*: 27-43.

Claramunt, C., Parent, C., and Thériault M. (1998). Design patterns for spatio-temporal processes. *Data Mining and Reverse Engineering*. Springer US, 1998. 455-475.

- Claramunt, C., et al. (2005). Real-time geographical information systems for its. Proceedings of the 5th IEEE International Conference on ITS Telecommunications.
- Claramunt, C., & Stewart, K. (2015). Special issue on spatio-temporal theories and models for environmental, urban and social sciences: where do we stand?.
- Claramunt, C., Ray, C., Camossi, E., Jusselme, A.-L., Hadzagic, M., Andrienko, G. L., Andrienko, N. V., Theodoridis, Y., Vouros, G. A., Salmon, L. (2017). Maritime data integration and analysis: recent progress and research challenges, EDBT 2017: 192-197.
- Clementini, E., Di Felice, P., Van Oosterom, P. (1993): A small Set of Formal Topological Relationships Suitable for End-User Interaction. In Proceedings of 3rd International Symposium on Advances in Spatial Databases, SSD '93, pp. 277-295. LNCS 692, Springer.
- Clementini, E., Di Felice, P. (1994): A Comparison of Methods for Representing Topological Relationships. *Information Sciences*, 80, pp. 1-34.
- Clementini, E., Di Felice, P. (1996): A Model for Representing Topological Relationships between Complex Geometric Features in Spatial Databases. *Information Sciences*, 90 (1-4): 121-136.
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1), 21-27.
- Dehn, M., Gartner, H., & Dikau, R. (2001). Principles of semantic modeling of landform structures. *Computers & Geosciences*, 27(8), 1005-1010
- Dodge, S., Weibel, R., & Lautenschütz, A. K. (2008). Towards a taxonomy of movement patterns. *Information visualization*, 7(3-4), 240-252.
- Douglas, K. & Douglas, S. (2003). *PostgreSQL: a comprehensive guide to building, programming, and administering PostgreSQL databases*. SAMS publishing, 2003.
- Egenhofer, M.J., Franzosa (1991): Point-Set Topological Spatial Relations. *International Journal of Geographic Information Systems*, 5(2): 161-174, 1991.

- Egenhofer, M.J., Clementini E., Di Felice P. (1994): Topological Relations between Regions with Holes. *International Journal of Geographoc Information Systems*, 8(2): 129-142, 1994.
- Erwig, M., Güting, R. H., Schneider, M. (1999) Spatio-temporal data types: An approach to modeling and querying moving objects in databases. *GeoInformatica* 3.3: 269-296.
- Etches, A., Claramunt, C., Bargiela, A., & Kosonen, I. (1998, April). An integrated temporal GIS model for traffic systems. In *GIS Research UK VI National Conference*.
- Fernandez, E. B., & Yuan, X. (2000, October). Semantic analysis patterns. In *International Conference on Conceptual Modeling* (pp. 183-195). Springer, Berlin, Heidelberg.
- Fileto, Renato, Alessandra Raffaetà, Alessandro Roncato. A Semantic Model for Movement Data Warehouses. *International Workshop on Data Warehousing and OLAP*. pp 47-56, 2014.
- Freksa, C., & Röhrig, R. (1993). Dimensions of qualitative spatial reasoning (pp. 483-492). GrKK.
- Galton, A.(1995) "Towards a qualitative theory of movement." *Spatial Information Theory A Theoretical Basis for GIS*. Springer Berlin Heidelberg. 377-396.
- Gemino, A., Wand, Y. (2005) Complexity and clarity in conceptual modeling: Comparison of mandatory and optional properties. *Data & Knowledge Engineering* 55: 301–326
- Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196), 779-782.
- Goodchild, Michael F. (2000) GIS and transportation: status and challenges. *GeoInformatica* 4.2: 127-139.
- Goodwin, R. A., Nestler, J. M., Anderson, J. J., Weber, L. J., & Loucks, D. P. (2006). Forecasting 3-D fish movement behavior using a Eulerian–Lagrangian–agent method (ELAM). *Ecological Modelling*, 192(1), 197-223.

- Güting, R. H., de Almeida, T., & Ding, Z. (2006). Modeling and querying moving objects in networks. *The VLDB Journal—The International Journal on Very Large Data Bases*, 15(2), 165-190.
- Güting, R. H., & Schneider, M. (1993) *Realms: A foundation for spatial data types in database systems. Advances in Spatial Databases. Springer Berlin Heidelberg.*
- Güting, R. H., Böhlen, M. H., Erwig, M., Lorentzos, N. A. (2000) *A Foundation for Representing and Querying Moving Objects. ACM Transactions on Database Systems.*
- Hägerstrand, T. (1970). What about people in regional science?. *Papers in regional science*, 24(1), 7-24.
- Hallot, P., Billen, R. (2007). Spatio-Temporal Configurations of Dynamics Points in a 1D Space. In: Gottfried, B. (ed.) *Behaviour Monitoring and Interpretation BMI 2007, Centre for Computing Technologies (TZI), University of Bremen, Germany, Osnabrück, Germany, pp. 77–90.*
- Hanson, S., & Huff, J. O. (1981). Assessing day-to-day variability in complex travel patterns. *Transportation Research Record*, 891, 18-24.
- Hanson, S., & Huff, J. O. (1985). Implications of analysis of multi-day data for spatial choice modelling. In *64th Annual Meeting of the Transportation Research Board, Washington, DC, January.*
- Hanson, S., & Huff, J. (1986). Classification issues in the analysis of complex travel behavior. *Transportation*, 13(3), 271-293.
- Hornsby, K., Egenhofer, M. J. (2000). Identity-Based Change: A Foundation For Spatio-Temporal Knowledge Representation. *International Journal of Geographical Information Science* 14(3): 207-224.
- Hornsby, K., Egenhofer, M. J. (2002). Modeling Moving Objects over Multiple Granularities. *Annals of Mathematics and Artificial Intelligence*: vol. 36, pp 177-194.
- Hung, C. C., & Peng, W. C. (2009, November). Clustering object moving patterns for prediction-based object tracking sensor networks. In *Proceedings of the 18th ACM conference on Information and knowledge management* (pp. 1633-1636).

ACM.

Jacobsen, J. D. (1968) Interactive graphics in data processing: geometric relationships for retrieval of geographic information. *IBM Systems Journal* 7.3-4: 331-341.

Jiang, S., Ferreira Jr, J., & Gonzalez, M. C. (2012, August). Discovering urban spatial-temporal structure from human activity patterns. In *Proceedings of the ACM SIGKDD international workshop on urban computing* (pp. 95-102). ACM.

Kao, R. C. (1963). The use of computers in the processing and analysis of geographic information. *Geographical Review*: 530-547.

Kitamura, R. (1988). An evaluation of activity-based travel analysis. *Transportation*, 15(1-2), 9-34.

Kuhn, W., (2005). Geospatial semantics: why, of what, and how?. In *Journal on data semantics III* (pp. 1-24). Springer Berlin Heidelberg.

Kung, C. H., & Soelberg A. (1986). Activity modeling and behavior modeling. Proc. of the IFIP WG 8.1 working conference on Information systems design methodologies: improving the practice. North-Holland Publishing Co..

Laube, P., van Kreveld, M., & Imfeld, S. (2005). Finding REMO—detecting relative motion patterns in geospatial lifelines. In *Developments in spatial data handling* (pp. 201-215). Springer Berlin Heidelberg.

Laube, P., Imfeld, S., & Weibel, R. (2005). Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science*, 19(6), 639-668.

Lee, W. C., & Krumm, J. (2011). Trajectory preprocessing. *Computing with spatial trajectories*. Springer New York. 3-33.

Li, Z., Ding, B. (2010). Mining periodic behaviors for moving objects. *ACM SIGKDD international conference on Knowledge discovery and data mining*: pp 1099-1108.

Liu, X., Yan, W. Y., Chow, J. Y. (2015). Time-geographic relationships between vector fields of activity patterns and transport systems. *Journal of Transport Geography* Vol42: pp. 22–33.

- Lbsyun Baidu. Retrieved 2016, from lbsyun.baidu, <http://lbsyun.baidu.com/index.php?title=wxjsapi/guide/getpoi>
- Matos, L., Moreira, J., & Carvalho, A. (2012). A spatiotemporal extension for dealing with moving objects with extent in Oracle 11g. *ACM SIGAPP Applied Computing Review* 12.2: 7-17.
- Merki, M., & Laube, P. (2012, April). Detecting reaction movement patterns in trajectory data. In Avignon, France: Proceedings of the AGILE 2012 International Conference on Geographic Information Science (pp. 25-27).
- Ménard, A., & Marceau, D. J. (2005). Exploration of spatial scale sensitivity in geographic cellular automata. *Environment and Planning B: Planning and Design*, 32(5), 693-714.
- Miller, H. J., & Bridwell S. A. (2009). A field-based theory for time geography. *Annals of the Association of American Geographers* 99.1: 49-75.
- Moreno, N., Wang, F., & Marceau, D.J. (2010). A geographic object-based approach in cellular automata modeling. *Photogrammetric Engineering & Remote Sensing* 76.2: 183-191.
- Nanni, M. (2013, September 30). *Mobility Data Mining*. Book chapter in *Mobility Data: Modeling, Management, and Understanding*, Cambridge University Press. ISBN-13: 978-1107021716.
- Noyon, V., Devogele, T., Claramunt, C. (2005) A formal model for representing point trajectories in two-dimensional spaces. *ER Workshops 2005*. LNCS, vol. 3770, pp. 208-217.
- Noyon, V., Claramunt, C. & Devogele, T. (2007). A Relative Representation of Trajectories in Geographical Spaces. *Geoinformatica*. 11:479–496.
- O Sullivan, David. (2005). Geographical information science: time changes everything. *Progress in Human Geography* 29.6: 749.
- Ong, R., Wachowicz, M., Nanni, M., & Renso, C. (2010, December). From pattern discovery to pattern interpretation in movement data. In 2010 IEEE International Conference on Data Mining Workshops (pp. 527-534). IEEE.

- Parent, C., Spaccapietra, S., & Zimányi, E. (2006). Conceptual modeling for traditional and spatio-temporal applications: the MADS approach. Springer Science & Business Media.
- Parent, C., Spaccapietra, S., Renso, C., Andrienko, G., Andrienko, N., Bogorny, V., ... & Theodoridis, Y. (2013). Semantic trajectories modeling and analysis. *ACM Computing Surveys (CSUR)*, 45(4), 42.
- Pas, E. I. (1988). Weekly travel-activity behavior. *Transportation*, 15(1-2), 89-109.
- Pentland, A., & Liu, A. (1999). Modeling and prediction of human behavior. *Neural computation*, 11(1), 229-242.
- Pfoser, D., Jensen, C.vS., Theodoridis, Y. (2000). Novel approaches to the indexing of moving object trajectories *Proceedings of VLDB*.
- Praing, R., Schneider, M. (2007) Modeling Historical and Future Spatio-temporal Relationships of Moving Objects in Databases. Volume 4802 of the series Lecture Notes in Computer Science: pp 318-327.
- Qian, D., & Yi, J. (2016). Hierarchical sliding mode control for under-actuated cranes (pp. 1-166). Springer, Berlin.
- Ramsey, P., (2004). PostGIS Introduction and Evaluation, Refrations Research Inc.
- Ramsey, P. (2016). PostGIS Manual. Refrations Research Inc.
- Raper, J., & Livingstone, D. (1995). Development of a geomorphological spatial model using object-oriented design. *International Journal of Geographical Information Systems* 9.4: 359-383.
- Renso, C., Baglioni, M., de Macedo, J. A. F., Trasarti, R., & Wachowicz, M. (2013). How you move reveals who you are: understanding human behavior by analyzing trajectory data. *Knowledge and information systems*, 37(2), 331-362.
- Rishe, N. (1992). Database design: the semantic modeling approach. McGraw-Hill, p. 528.
- Roger, W., Engelen, G., & Uljee, I. (1997) The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and*

planning B 24: 323-344.

Schlich, R., & Axhausen, K. W. (2003). Habitual travel behaviour: evidence from a six-week travel diary. *Transportation*, 30(1), 13-36.

Sideridis, S., Pelekis, N., & Theodoridis, Y. (2016). On querying and mining semantic-aware mobility timelines. *International Journal of Data Science and Analytics*, 2(1-2), 29-44.

Spaccapietra, S., Parent, C., Damiani, M. L., de Macedo, J. A., Porto, F., & Vangenot, C. (2008). A conceptual view on trajectories. *Data & knowledge engineering*, 65(1), 126-146.

Srivastava, G., & Schönfelder, S. (2003). On the temporal variation of human activity spaces. Zurich: ETH, Eidgenössische Technische Hochschule Zürich, Institut für Verkehrsplanung und Transportsysteme.

Stell, J., Del Mondo, G., Thibaud, R., & Claramunt, C. (2011, September). Spatio-temporal evolution as bigraph dynamics. In *International Conference on Spatial Information Theory* (pp. 148-167). Springer Berlin Heidelberg.

Stewart, K., Fan, J. & White, E. (December 2013) Thinking about Space-Time Connections: Spatiotemporal Scheduling of Individual Activities. *Transactions in GIS.V 17*, pp: 791–807.

Story, P. A., & Worboys, M. F. (1995). A design support environment for spatio-temporal database applications. *Spatial Information Theory A Theoretical Basis for GIS*. Springer Berlin Heidelberg. 413-430.

Stuchenschmidt, H., Parent, C., & Spaccapietra, S. (Eds.). (2009). *Modular ontologies: concepts, theories and techniques for knowledge modularization* (Vol. 5445). Springer.

Su, J., Xu, H., Ibarra, O. H. (2001). *Moving Objects: Logical Relationships and Queries*. Volume 2121 of the series *Lecture Notes in Computer Science* pp 3-19.

Susilo, Y., & Kitamura, R. (2005). Analysis of day-to-day variability in an individual's action space: exploration of 6-week Mobidrive travel diary data. *Transportation Research Record: Journal of the Transportation Research Board*, (1902), 124-133.

- Susilo, Y. O., & Axhausen, K. W. (2007). How firm are you? A study of the stability of individual activity-travel-location patterns using the Herfindahl-Hirschman Index. In 11th World Conference on Transport Research.
- Thériault, M. and Claramunt, C., 1999, La modélisation du temps et des processus dans les SIG: Un moyen d'intégration pour la recherche interdisciplinaire, *Revue Internationale de Geomatique*, Ed. Hermes, Paris, France, 9(1), pp. 67-103.
- Thériault, M., Claramunt, C. & Villeneuve, P., 1999, A spatio-temporal taxonomy for the representation of spatial set behaviours, in *Spatio-temporal Database Management*, M. Böhlen, C. Jensen, and M. Scholl (eds.), Springer-Verlag, LNCS 1678, pp. 1-19
- Thériault, M., Claramunt, C., Séguin, A., & Villeneuve, P. (2002). Temporal GIS and statistical modelling of personal lifelines." In *Advances in Spatial Data Handling*, pp. 433-449. Springer Berlin Heidelberg.
- Trivedi, S. K., Dey, S., Kumar, A., & Panda, T. K., *Handbook of Research on Advanced Data Mining Techniques and Applications for Business Intelligence*. IGI Global, Feb 14, 2017
- Torrens, P. M., & Benenson, I. (2005). Geographic automata systems. *International Journal of Geographical Information Science* 19.4: 385-412.
- Valsecchi, P., Claramunt, C., & Peytchev, E. (1999). OSIRIS: an inter-operable system for the integration of real-time traffic data within GIS. *Computers, Environment and Urban Systems*, 23(4), 245-257.
- Vazirgiannis, M., Wolfson, O. (2001). A Spatiotemporal Model and Language for Moving Objects on Road Networks. Chapter *Advances in Spatial and Temporal Databases*, Volume 2121 of the series *Lecture Notes in Computer Science*: pp 20-35.
- Wakamiya, S., Belouaer, L., Brosset, D., Kawai, Y., Lee, R., Sumiya, K. and Claramunt, C., 2015, Measuring crowd mood in city space through Twitter. In *Proceedings of the 14th International Symposium on Web and Wireless GIS (W2GIS 2015)*, Springer LNCS, pp. 37-42.
- Wu, F. (2002). Calibration of stochastic cellular automata: the application to

- rural-urban land conversions. *geographical information science*, vol. 16, no. 8, 795–818.
- Wu, J., Claramunt, C., & Deng, M. (2014, November). Modelling movement patterns using topological relations between a directed line and a region. In *Proceedings of the 5th ACM SIGSPATIAL International Workshop on GeoStreaming* (pp. 43-52). ACM.
- Xu, J. & Güting, R. H. (2013). A generic data model for moving objects. *Geoinformatica*, vol. 17, pp 125-172, 2013
- Yan, Z., Macedo, J., Parent, C., & Spaccapietra, S. (2008). Trajectory ontologies and queries. *Transactions in GIS*, 12(s1), 75-91.
- Yan, Z., Chakraborty, D., Parent, C., Spaccapietra, S., & Aberer, K. (2013). Semantic trajectories: Mobility data computation and annotation. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 4(3), 49.
- Zhao, L., Jin, P., Zhang, L., Wang, H., & Lin, S. (2011, April). Developing an oracle-based spatio-temporal information management system. In *International Conference on Database Systems for Advanced Applications* (pp. 168-176). Springer Berlin Heidelberg.
- Zheng, Y., Zhang, L., Xie, X., & Ma, W. Y. (2009, April). Mining interesting locations and travel sequences from GPS trajectories. In *Proceedings of the 18th international conference on World wide web* (pp. 791-800). ACM.
- Zheng, Y., Li, Q., Chen, Y., Xie, X., & Ma, W. Y. (2008, September). Understanding mobility based on GPS data. In *Proceedings of the 10th international conference on Ubiquitous computing* (pp. 312-321). ACM.
- Zheng, Y., Xie, X., & Ma, W. Y. (2010). GeoLife: A Collaborative Social Networking Service among User, location and trajectory. Invited paper, in *IEEE Data Engineering Bulletin*. 33, 2, pp. 32-40.
- Zheni, D., Frihida, A., Ghezala, H. B., & Claramunt, C. (2009). A Semantic Approach for the Modeling of Trajectories in Space and Time. Volume 5833 of the series *Lecture Notes in Computer Science*: pp 347-356.
- Zheni, D., Frihida, A., Claramunt, C. & Ben Ghézala, H. H.(2015). A semantic-based

data model for the manipulation of trajectories: Application to urban transportation. In Proceedings of the 14th International Symposium on Web and Wireless GIS (W2GIS 2015), Springer LNCS, pp. 124-142

Un modèle spatio-temporel sémantique pour la modélisation de mobilités en milieu urbain

Résumé—La croissance rapide et la complexité de nombreuses villes contemporaines offrent de nombreux défis de recherche pour les scientifiques à la recherche d'une meilleure compréhension des mobilités qui se produisent dans l'espace et dans le temps. À l'heure où de très grandes séries de données de trajectoires en milieu urbain sont disponibles grâce à profusion de nombreux capteurs de positionnement et de services de nombreuses et nouvelles opportunités de recherche et d'application nous sont offertes. Cependant, une bonne intégration de ces données de mobilité nécessite encore l'élaboration de cadres méthodologiques et conceptuels tout comme la mise en œuvre de bases de données spatio-temporelles qui offriront les capacités appropriées de représentation et de manipulation des données. La recherche développée dans cette thèse introduit une modélisation conceptuelle et une approche de gestion de base de données spatio-temporelles pour représenter et analyser des trajectoires humaines dans des espaces urbains. Le modèle considère les dimensions spatiales, temporelles et sémantiques afin de tenir compte de l'ensemble des propriétés issues des informations de mobilité. Plusieurs abstractions de données de mobilité et des outils de manipulation de données sont développés et expérimentés à partir d'une large base de données de trajectoires disponibles dans la ville de Pékin. L'intérêt de l'approche est double: il montre d'une part que de larges ensembles de données de mobilité peuvent être intégrés au sein de SGBD spatio-temporels extensibles; d'autre part des outils de manipulation et d'interrogation spécifiques peuvent être dérivés à partir de fonctions intégrées au sein d'un langage d'interrogation. Le potentiel de l'approche est illustré par une série d'interrogations qui montrent comment à partir d'une large base de données de trajectoires quelques patrons de déplacements peuvent être obtenus.

Mots-clefs—Trajectoires urbaines, Modélisation spatio-temporelle, Bases de données spatio-temporelles

A Conceptual and Semantic Modelling Approach for the Representation and Exploration of Human Trajectories

Abstract—Massive trajectory datasets generated in modern cities generate not only novel research opportunities but also important methodological challenges for academics and decision-makers searching for a better understanding of travel patterns in space and time. This PhD research is oriented towards the conceptual and GIS-based modeling of human displacements derived from large sets of urban trajectories. The motivation behind this study originates from the necessity to search for and explore travel patterns that emerge from citizens acting in the city. Our research introduces a conceptual modelling framework whose objective is to integrate and analyze human displacements within a GIS-based practical solution. The framework combines conceptual and logical models that represent travel trajectories of citizens moving in a given city. The whole approach has been implemented in a geographical database system, experimented in the context of transportation data, and enriched by a series of query interface manipulations and specific functions that illustrate the potential of our whole framework for urban studies. The whole framework has been experimented on top of the Geolife project and large trajectories datasets available in the city of Beijing. Overall, the findings are twofold: first, it appears that our modelling framework can appropriately act as an extensible geographical database support for the integration of large trajectory datasets; second the approach shows that several emerging human displacements can be explored from the manipulation of large urban trajectories.

Keywords—Urban trajectories; Spatio-temporal data modeling; Spatio-temporal database