

NANYANG
TECHNOLOGICAL
UNIVERSITY

Visual Analytics for Massive Urban Public Transport Data

ZENG Wei

School of Computer Engineering

2015

Visual Analytics for Massive Urban Public Transport Data

ZENG Wei

School of Computer Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2015

Abstract

Public transport systems (PTSs) play an important role in modern cities. From the perspective of commuters, PTS provides shared and rapid transport services that are essential for the general public in a city. From the perspective of city management and urban planning, PTS has significant economical, social and environmental impacts to an entire city. Hence, studying PTS is highly beneficial to both individuals as well as a city as a whole, and it has long been a hot topic in transport research. However, many conventional transport researches have long been relying on simulation and survey data, making the results less of conviction.

Thanks to recent advances in sensing technologies, such as RFID cards, laser scanners and GPS devices, movements acquisition has become convenient. Vast amount of urban public transport data has been collected automatically and pervasively, promoting more research focus on analyzing and exploring public transport data when studying PTS. However, analyzing massive urban public transport data is a challenging task due to its high-complex, large-size, multi-mode and spatio-temporal characteristics. To get over these challenges, visual analytics show great potential as they can make the way of processing public transport data transparent: Visual analytics can provide interactive means for transport researchers to examine the actual processes of analyzing data instead of just the results.

This thesis investigates advanced visualization technologies for analyzing and exploring massive urban public transport data that consists of commuter RFID card data, transport network and transit schedule in Singapore. To address various analytical tasks raised by transport researchers, a family of novel visual analytics systems have been developed. Specifically, three aspects of high-level information, which are essential in transport modeling and analysis processes, have been extracted from the input dataset for visualization and exploration:

- *Interchange Pattern*, which describes how moving objects redistribute when entering and passing through a junction node in a traffic network. A novel visual representation, namely the interchange circos diagram, has been proposed to present interchange pattern emerged from the public transport data.
- *Waypoints-Constrained OD Pattern*, which restricts origin-destination (OD) pattern with commuter trajectories passing through user-specified entry and exit waypoints in a transport network. A novel unified visual representation, namely the waypoints-constrained OD view, has been proposed to explore waypoints-constrained OD pattern.
- *Mobility*, which can be considered as the travel efficiencies of commuters through PTS. An integrated visualization with three modules: isochrone map view, isotime flow map view and OD-pair journey view, has been proposed to address a family of analytical tasks based on inputs from transport researchers.

Acknowledgments

It would not have been possible to write this doctoral thesis without the help and support of the kind people around me, to only some of whom it is possible to give particular mention here.

Above all, I would like to thank my supervisors Prof. Chi-Wing Fu and Prof. Stefan Müller Arisona for their continuous support of my Ph.D. study and research. This thesis would not have been possible without their invaluable guidance. I would also like to thank Prof. Huamin Qu, an expert in the fields of visualization and visual analytics, for his precious advice on various projects. Besides, I would specially thank Prof. Ian Vince McLoughlin, who led me into research and helped a lot in the beginning. It has been a great time working with them, for their valuable research skills and personal fascination that I have learned.

Meanwhile, many thanks to my colleagues at Future Cities Laboratory, specially to members of Module IX Simulation Platform and Module VIII Mobility and Transportation Planning, for their kind support and assistance. In addition, I would also like to acknowledge my colleagues at NTU GameLab for their assistance and constant support on many research projects.

Last but not least, I would like to thank my family for their constant supports and encouragement. Specially, I would like to thank my wife Ms. Shanshan Chen, who make me always feel warm and sweet with her ultimate care and love.

Contents

Abstract	ii
Acknowledgments	iv
List of Figures	ix
1 Introduction	1
1.1 Background	1
1.2 Research Objectives	2
1.3 Research Contributions	5
1.4 Thesis Organization	9
2 Related Work	10
2.1 Movement Pattern Analysis	10
2.1.1 Movement Pattern Mining	11
2.1.2 Movement Pattern Applications	12
2.2 Visual Analytics for Movement Data	13
2.2.1 Temporal-Oriented Visual Analytics	13
2.2.2 Spatial-Oriented Visual Analytics	15
2.2.3 Spatio-Temporal-Oriented Visual Analytics	16
2.2.4 Pattern-Oriented Visual Analytics	18
2.3 Visualizing Urban Transport	20
2.3.1 Transport Network Visualization	20
2.3.2 Urban Traffic Simulation	21
2.3.3 Urban Traffic Visual Analytics	22
2.4 Inadequacy of Existing Methods	23

3	Massive Urban Public Transport Data	25
3.1	Transport Terminologies	25
3.2	Input Public Transport Dataset	26
3.2.1	EZLink Data	26
3.2.2	Public Transport Network	28
3.2.3	Transit Schedule	29
3.3	Data Characteristics	29
3.4	Data Applications	32
4	Visual Analytics for Interchange Patterns	35
4.1	Introduction	35
4.2	Overview	37
4.2.1	Formal Definition: Interchange	37
4.2.2	System Workflow	38
4.3	Interchange Circos Diagram	39
4.3.1	Circos Figures	40
4.3.2	Initial Design: Interchange Circos Diagram	41
4.3.3	Improving Our Visual Design	42
4.3.4	Coloring Arc Elements	44
4.3.5	Positioning Arc Elements	45
4.4	Interface: Visualizing Interchange	46
4.4.1	Multi-scale Visualization	46
4.4.2	Interaction	47
4.5	Implementation and Results	48
4.5.1	Implementation	48
4.5.2	Case Study 4-1: Interchange at Metro Stations	49
4.5.3	Case Study 4-2: Intersection Capacity Utilization	51
4.6	Discussion	52
5	Visual Analytics for Waypoints-Constrained OD Patterns	54
5.1	Introduction	55
5.2	Overview	56

5.3	Waypoints-Constrained OD Query	59
5.3.1	Interactive Waypoints Specification	59
5.3.2	Hashing-Based Trajectory Query	59
5.3.3	Map View	61
5.4	Waypoints-Constrained OD View	62
5.4.1	Design Philosophy	63
5.4.2	The In-Flow and Out-Flow Views	65
5.4.3	The OD-Flow Temporal View	66
5.4.4	Discussion: Design Alternatives	68
5.4.5	Ordering of Origin/Destination Boxes	70
5.4.6	User Interaction	71
5.5	Evaluation and Discussion	71
5.5.1	Performance Evaluation: Trajectory Query	71
5.5.2	Study 1: Transportation Network Usage Analysis	72
5.5.3	Study 2: Daily Pendulum Movements Exploration	74
5.5.4	Expert Interview	77
5.5.5	Discussion	78
6	Visual Analytics for PTS Mobility	80
6.1	Introduction	81
6.2	Overview	83
6.2.1	Problem Definition	83
6.2.2	Analytical Tasks and Mobility-Related Factors	84
6.2.3	System Workflow	85
6.3	Modeling PTS Mobility	86
6.3.1	PTS Mobility Model	86
6.3.2	Estimating Mobility-Related Factors	87
6.4	Visualization Design	89
6.4.1	Isochrone Map View	90
6.4.2	Isotime Flow Map View	91
6.4.3	OD-pair Journey View	94

6.4.4	Design Alternatives	96
6.5	Evaluation	98
6.5.1	Case Study 6-1: Spatial Variation of PTS Mobility	98
6.5.2	Case Study 6-2: Analyzing Mobility-Related Factors	101
6.5.3	Experts Interview	102
6.6	Discussion	104
7	Conclusion and Future Work	106
7.1	Conclusion	106
7.2	Contributions	108
7.3	Future Work	109
	References	111
	Publications	127

List of Figures

1.1	An example of the EZLink card (left) and the card reader (right) in Singapore. .	2
1.2	The visual analytics process proposed by Kiem et al. [71].	3
1.3	The information visualization reference model proposed by Card et al. [28]. . .	4
1.4	Multi-scale analytical tasks of OD patterns emerged from transport domain. 1) Small-scale: analyze how moving objects redistribute themselves at junctions, e.g., crossroads, bus stops; 2) Medium-scale: study origins and destinations of the trajectories successively passing through an entry and exit waypoints; and 3) Large-scale: explore all reachable destinations from a starting location within certain time period.	6
1.5	Novel visual representations proposed in this thesis from left to right: inter- change circos diagram [151], waypoints-constrained OD view [152] and OD trip journey view [150], corresponding to the small-, medium- and large-scale OD-related analytical tasks, respectively.	8
2.1	Examples of 3D visualizations of movement data: (a) GeoTime [69] and (b) stacking-based visualization [123].	17
2.2	Examples of edge bundling techniques applied to visualize US migrations: (a) geometry-based [34], (b) winding roads [76], (c) force-directed [60], and (d) kernel density estimation-based [62].	19
2.3	An overview of the visual analysis system to explore traffic jam based on urban traffic data proposed by Wang et al. [135].	23
2.4	Minima events identification in NYC using topological analysis in urban traffic data proposed by Doraiswamy et al. [37].	23

3.1	An example of the EZLink records revealing a child/student journey traveling to NTU (stop 27321) in the afternoon.	28
3.2	Singapore public transport network in 2011. Thin brown lines indicate the bus network, while thick green, red, purple, and yellow lines indicate the XX, YY, ZZ and WW MRT lines, respectively.	29
3.3	Road network travel efficiency in the period of 08:00 - 08:15. Green and red colors indicate high and low travel efficiency w.r.t. mean, respectively.	31
3.4	Temporal variation of alighting commuter volume distribution at XXX MRT station during different periods. Left: 08:00 - 09:00; Right: 11:00 - 12:00. . . .	31
3.5	Origin and destination density in the morning retrieved from the input dataset, with red color indicates more commuters originating in and blue color indicates more commuters destined, respectively.	33
3.6	Multiple routes retrieved in-between red circle as the origin and blue circle as the destination from the input dataset.	34
4.1	The interchange information (ten trajectories) at this junction can be summarized as a 5-by-5 matrix.	38
4.2	System workflow: from (a) a set of raw trajectory paths, to (b) traffic networks of different spatial scales, (c) interchange statistics, (d) interchange circos diagrams per junction node, and (e) our interchange visualization with user interaction.	38
4.3	Example circos figures developed by Krzywinski et al. [74] for examining the mutual relationship among genomes.	40
4.4	An initial design of interchange circos diagram with arc elements representing the exterior links and junction itself, and ribbons connecting arc elements representing the flows.	41
4.5	Developing the interchange circos diagram from the original circos figure: (a) the initial design in Figure 4.4; (b) use a grey-colored flyover ring (like a source/sink) for the junction itself; (c) bundle pairs of bi-directional ribbons to reduce the visual cluttering; and (d) draw white and black curved statistics boxes to present the total outgoing and incoming flow volumes.	42

4.6	Interchange Circos Diagrams in different scales: (a) city scale; (b) regional scale; and (c) road network scale.	47
4.7	Interchange patterns at different train stations (<i>a-d</i>) in the Singapore Metro system.	49
4.8	Exploring the temporal changes (over a day) in the interchange patterns at four different train stations (<i>a-d</i>) in the Singapore MRT system.	50
4.9	Comparison of lower (left) and higher (right) ICU ratings at a road junction during different time periods. (a) The traffic flow from yellow to violet dominates the junction utilization; moreover, both the orange and yellow connecting links are highly unbalanced. (b) Traffic flows from different links in the junction are fairly balanced and the incoming/outgoing flows for each connecting links are also fairly balanced.	52
4.10	Comparing interchange circos diagrams with existing visualization approach. (a) Two sets of raw trajectories; (b) Existing approach aggregates flows between pairs of locations and draws arrows to indicate the aggregated flow volume; (c) Our interchange circos diagrams are able to reveal the detail on the interchange patterns.	53
5.1	Overview of our waypoints-constrained OD visual analytics interface. We can interactively (a) select and modify the entry (red) and exit (blue) waypoints in the transportation network, (b) filter passenger trajectories that successively pass through the selected waypoints, and (c) present the waypoints-constrained OD view to support the analytical tasks.	56
5.2	Interactive waypoints specification: a) mode 1: two successive clicks to select an entry and then an exit waypoint; b) mode 2: drag to define a path from the entry waypoint to exit waypoint; c) mode 3: click to select an entry waypoint and an outgoing direction (red arrow); and d) click to select and then drag to modify an existing waypoint.	58
5.3	The map view overviews the waypoints-constrained trajectories: a) an origin; b) red and blue glyphs as entry and exit waypoints, resp.; and c) a destination.	61
5.4	An example <i>waypoints-constrained OD view</i> with three components: (a) in-flow view, (b) OD-flow temporal view, and (c) out-flow view, as an integrated and coordinated solution to support the visual analytical tasks.	62

5.5	An example in-flow view: vertical boxes show the origins with heights to indicate flow volumes, and ribbons to show flow aggregation before the entry waypoint.	64
5.6	An example OD-flow temporal view that presents the OD patterns for the trajectories shown in Figure 5.3: (a) the two vertical bars represent the origin waypoint (left) and destination waypoint (right), and the colored boxes along the two bars represent the corresponding origins and destinations, respectively, (b) ribbons connect the origin and destination boxes and show the OD flow volumes in-between the corresponding OD pairs with embedded heat maps to reveal flow volume variation over time, and (c) relative (average) travel distance of each OD pair. Note also that the heat maps are colored by the color bar on top, e.g., the rightmost deep red is used for flow volume in range (216, 720], etc.	65
5.7	Design alternative: small multiples that show flow volume difference over six different time periods with the same entry and exit waypoints as in Figure 5.4. The individual diagrams are too small to present useful information.	68
5.8	Illustration: effects of ordering origin and destination boxes by flow volumes (left) and by overlapping minimization (right) in the OD-flow temporal view.	70
5.9	Case study 1: transportation network usage analysis. (a) & (b) present the morning and evening OD-flow temporal views, respectively, for the same pair of waypoints shown in the map view (c). Highlighted ribbon (d) shows that trajectories leaving WW27 station for XX-line stations (red) always pass through the WW branch (green) instead of the YY branch (purple) (note: they are two different time-efficient paths between the waypoints, see again the map view), while (e) shows that trajectories transferred from YY1 station to XX stations always pass through the YY branch.	73
5.10	Case study 2: daily pendulum movement exploration. (a) & (b) present an interesting pendulum movement pattern, which illustrates home-to-work movements through the red arrow in (c) in the morning (06:00-10:00), and work-to-home movements through the blue arrow in (c) in the evening (16:00-20:00).	75

6.1	Overview of our system workflow. In the data preprocessing stage, we integrate the input data to estimate various mobility-related factors. These information are then presented in the visual exploration and analysis stage, which comprises three visualization modules: (i) isochrone map view, (ii) isotime flow map view, and (iii) OD-pair journey view, which complement one another and work together to support various analytical tasks.	85
6.2	The isochrone map view presents spatial-reachability regions from 08:00 at origin A (red icon), which is station XX16/WW3 in the city center, by contour lines and areas over the geographical map. Dark blue and light blue indicate [0, 30] and (30, 60] minutes, respectively.	90
6.3	The example isotime flow map view above shows time-efficient journeys started from the city center station XX16/WW3 at 08:00 within a duration T of one hour. These journeys are arranged in a parallel isotime fashion with corresponding travel time and travel efficiency revealed by the horizontal time axis and color code on nodes (see the color map on the right), respectively.	92
6.4	Left: A tree structure that includes all time-efficient journeys from starting node A. Right: the result of applying our spatial layout algorithm to arrange the nodes in the visualization view.	93
6.5	Left: up-and-down wobbling. Right: we fix it by vertically shifting the nodes by node-based moving-window averaging.	93
6.6	Branch routing. Left: branches overlap issue may occur. Right: we resolve it by horizontally shifting the Bézier control points.	94
6.7	The OD-pair journey view focuses on Tasks 6-4 & 6-5, presenting detailed mobility-related information along routes from A to user-specified nodes (left), as well as their round-the-clock variations using our proposed visual representation: <i>mobility wheel</i> (right). Note also that we employ the standard colors of Singapore MRT lines to show the riding time on MRT, and encode bus lines by yellow.	95
6.8	Designs alternative that we have explored for handling Tasks 6-2 & 6-3: (a) time-scaled network distortion and (b) radial isotime layout.	96
6.9	Case Study 6-1(a): Exploring the spatial variation of PTS mobility from <i>MRT station “XX24/YY1”</i> at 08:00 in the morning.	99

6.10 Case Study 6-1(b): Exploring the spatial variation of PTS mobility from a rural-area bus stop at 08:00 in the morning.	100
6.11 Case Study 6-2: Analyzing detailed mobility-related factors along two different user-specified journeys starting from <i>MRT station “XX24/YY1”</i> to destinations <i>Bus stop 67852</i> and <i>MRT station “WW13/ZZ12.”</i> (a) Estimated waiting, riding and transfer time can be clearly presented along the pathways; we can see together the related MRT and bus trips, as well as the transfer points. (b & c) show the zoomed views of the mobility wheels in (a) for presenting the round-the-clock mobility patterns.	101

Chapter 1

Introduction

1.1 Background

Public transport system (PTS) is an important infrastructure in most modern cities. From the perspective of city management and urban planning, PTS is more than a service provider. It has several significant impacts to a city: *economically*, since PTS reduces the overall transport cost of a city [24, 81]; *socially*, since PTS ensures all members of a city are able to travel [128]; and *environmentally*, since PTS generally saves more energy than private transport [17]. From the perspective of commuters, PTS provides not only shared commuting services available to everyone in the general public, but also rapid transit services via trains and subways, thus being capable of efficiently moving large volume of people across a city. This is particularly important for big cities in Asia, where private cars and taxis are not the major modes of transport, and most people rely on PTS to travel in their daily lives.

Given the importance of PTS, many transport researches have been carried out to improve traffic management and transport planning. In general, these researches need to possess an accurate picture of the underlying movement patterns [117]. However, the majority of empirical transport researches on understanding human movements rely on travel surveys [27] that have obvious limitations, including high expense, small samples and low update frequency [117].

Thanks to the recent development of sensing devices, such as personalized radio-frequency identification (RFID) cards, GPS devices, laser scanners and video recorders, massive amounts of urban public transport data have been collected in an automatic and pervasive way, providing transport researchers a superior alternative to study the movement patterns.



Figure 1.1: An example of the EZLink card (left) and the card reader (right) in Singapore.

Specially in Singapore, commuters tap-in/out their personalized RFID card (referred as EZLink card as shown in Figure 1.1(left)) to enter/exit the PTS that includes a bus system and mass rapid transit (MRT). The EZLink card reader (Figure 1.1(right)) will record information like the boarding and alighting stops of each commuter trip, see Section 3.2 for more details, and the data is saved and owned by Singapore Land Transport Authority (LTA). LTA shares the data and cooperates with other departments like Urban Redevelopment Authority (URA) and research institutes like Future Cities Laboratory (FCL) to study and analyze how commuters utilize the public transport services for the purpose of building up and improving the land transport system in Singapore [15]. This thesis is a collaborative work between LTA and FCL, targeting at investigating the commuter flows of Singapore PTS by exploring the massive EZLink data such that to manage, plan and optimize these flows.

1.2 Research Objectives

Although the EZLink data contains plentiful information, the data itself has no value; instead, only the right information extracted from the raw data shows its value in the process of de-

signing strategies and making decisions. To automatically mine useful information from the data, numerous data analysis methods such as data mining and statistics can be employed. However, these methods still face significant challenges such as algorithm scalability and data heterogeneity [116]. More importantly, these approaches have the problem of understanding and analyzing the analyses: The results are only reliable in well-defined and well-understood problems [70, 71].

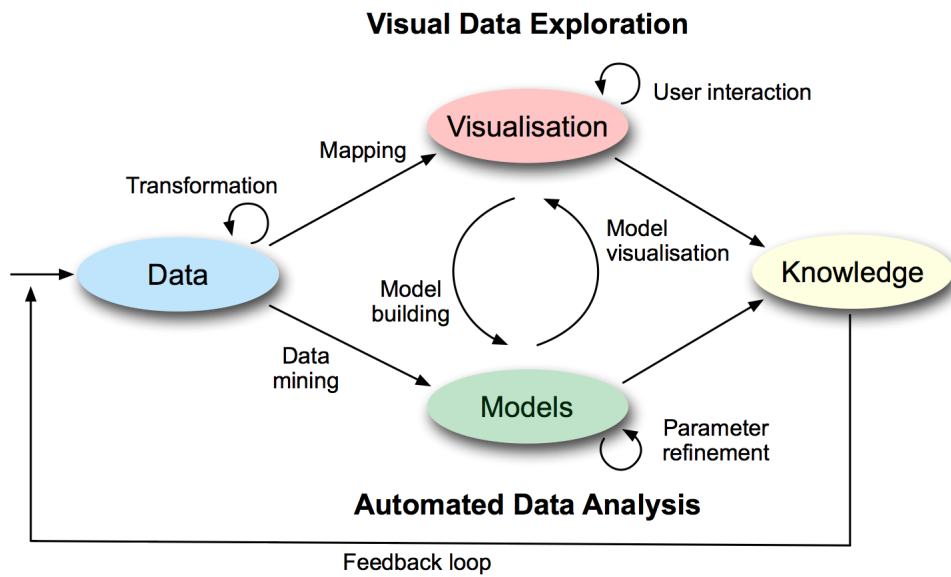


Figure 1.2: The visual analytics process proposed by Kiem et al. [71].

Visual analytics has been developed to help address these challenges in recent years, which is “*the science of analytical reasoning facilitated by interactive visual interfaces*” [118]. Figure 1.2 presents a popular visual analytics process for data exploration proposed by Kiem et al. [71]. The illustration reveals two important features of visual analytics. First, visual analytics “*combines automated analysis with interactive visualizations*” [71], which allows humans to progressively refine and evaluate the analysis results [116]. This is important: Analysts can examine the analysis procedures if the information extracted is suspicious, and thus appropriate designed visual analytics can make the way of data mining and knowledge discovery transparent [118]. Second, visual analytics is an inter-disciplinary research field that combines

various related research domains such as visualization, data mining, human-computer interaction, and application domains, such as transport and GIS. Existing procedures and models in application domains can contribute to the development of effective visual analytics tools. On the other hand, effective visual analytics tools can also help domain experts refine their models and analysis process.

However, there is very little amount of work to develop visual analytics for transport researchers to explore and analyze massive urban public transport data. This is particularly due to the special characteristics of the data, including 1) *big data volume*: the data consists of millions of commuter trips each day; 2) *complex network*: the public transport includes around hundred MRT stations and thousands bus stops; 3) *multi-modes transport*: commuters can travel through the public transport in multiple modes such as bus and MRT, and they can also switch between these modes by walking; and 4) *spatial and temporal patterns*: the movement patterns can vary drastically in both time and space dimensions. See Section 3.3 for a more detailed description of characteristics of the input dataset.

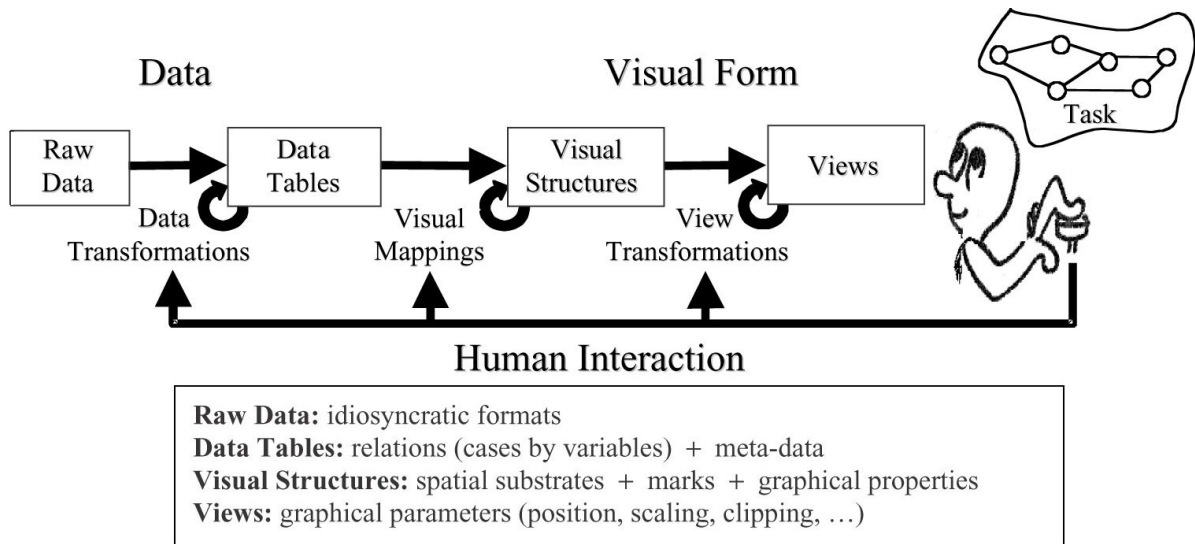


Figure 1.3: The information visualization reference model proposed by Card et al. [28].

Therefore, we set the overall objective of this thesis *to develop effective visual analytics systems*

for transport researchers to facilitate the understanding of commuter flows from the massive urban public transport data. To achieve this goal, we closely follow the agenda in the *information visualization reference model* [28] as shown in Figure 1.3. This model describes the process of developing interactive visualizations motivated by a user’s task in three steps: 1) transform the raw data into data tables; 2) map the data tables into visual structures; and 3) build visualizations from the visual structures. The user can observe the views, and interactively control the process by adjusting any of the steps, if necessary.

Based on this model, we can explicitly subdivide the overall objective to the following specific goals in the course of this thesis:

- **Data Transformation:** Identify the analytical tasks that transport researchers would like to address, and develop efficient mining techniques to extract the related information from the raw EZLink data.
- **Visual Mapping:** Explore state-of-the-art visualization techniques that can be used to present the extracted information, and formulate novel visual encodings to map the information into more efficient visual representations.
- **View Transformation:** Develop interactive visual analytics systems that employ the visual encodings and interaction techniques to meet transport researchers’ needs of analyzing and exploring the information.

1.3 Research Contributions

In close collaboration with transport researchers from the Mobility and Transport Planning Module at Future Cities Laboratory, we identified a key focus in transport domain, i.e., origin-destination (OD) pattern. As it summarizes people and goods movements across transport

zones, OD pattern provides primary information in transport planning and forecasting process [129]. Although many visualization methods can be employed to visualize OD patterns, including flow maps [120, 121], edge-bundling [34, 60, 76, 39, 62] and matrix visualizations [46, 51], we find they cannot fully meet needs of the transport researchers.

Generally speaking, these visualizations are dedicated to presenting global movements across a whole area or region. However, in many cases, transport researchers would like to explore local features of OD patterns, such as to study the trips passing through a specific location or path, and to analyze not only the movements, but also OD-related travel time information. As shown in Figure 1.4, we summarize these OD-related analytical tasks emerged from transport domain into the following three scales:

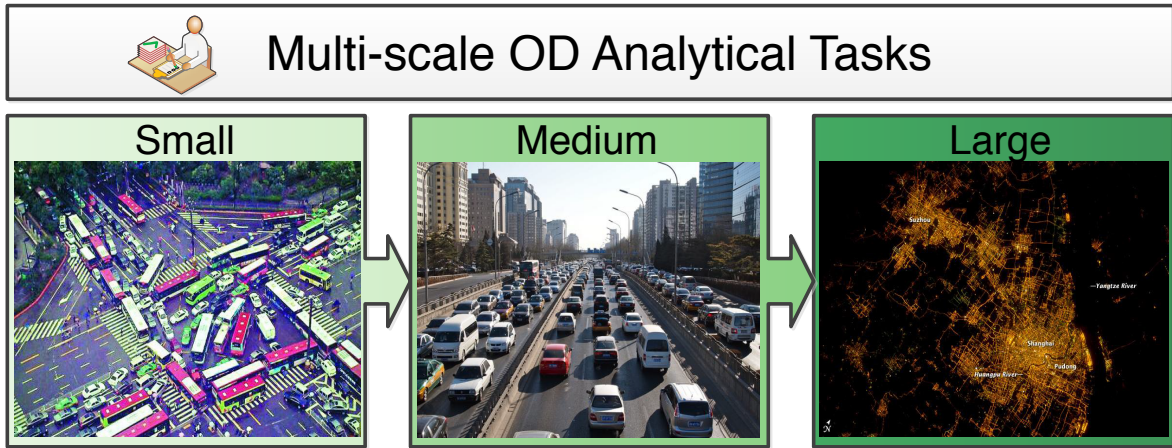


Figure 1.4: Multi-scale analytical tasks of OD patterns emerged from transport domain. 1) Small-scale: analyze how moving objects redistribute themselves at junctions, e.g., crossroads, bus stops; 2) Medium-scale: study origins and destinations of the trajectories successively passing through an entry and exit waypoints; and 3) Large-scale: explore all reachable destinations from a starting location within certain time period.

- Small-scale (Chapter 4)

Small-scale OD pattern, which is referred as interchange pattern in this thesis, describes how moving objects redistribute when entering/passing through a junction node in the transport network. Interchange pattern is a highly valuable means not only for unveiling

human movements at junctions, such as crossroads and bus stops, but also for assisting transport planning. For instance, interchange information can help reveal the road junction utilization and suggest crossroad design, e.g., adding a fork.

- Medium-scale (Chapter 5)

Medium-scale OD pattern, which is referred as waypoints-constrained OD pattern in this thesis, associates with trajectories that successively pass through an entry and exit waypoints in the transport network. Such aspect has not been explored in previous visualization research, and has practical value in transport, e.g., the study [132] showed that only a few (less than 2%) of the road segments in urban areas give rise to congestions.

- Large-scale (Chapter 6)

Large-scale OD pattern explores all reachable destinations from a starting location within certain time period, which considers the mobility in the entire transport system. Mobility measures a transport system's ability to move goods and people to their destinations based on the quantity and quality of physical travel [82]. Traditional transport planning aims at improving the mobility of transport systems. Thus, studying the mobility of a transport system is highly beneficial to both individuals as well as an entire city as a whole.

This thesis has made several contributions to develop novel visual analytics for massive urban public transport data. In more detail, we have accomplished the following achievements to meet the objectives of this thesis:

- **Data Transformation**

Multi-scale analytical tasks have been identified from transport domain, and corresponding information have been successfully transformed from the massive urban public transport data: Chapter 4 studies interchange pattern, where the data is transformed into interchange matrices, summarizing the movements at junctions; Chapter 5 focuses on

waypoints-constrained OD pattern, where an efficient hashing-based query method has been proposed to extract related information from the data in real-time; and Chapter 6 explores mobility of a PTS, where detailed mobility-related factors, such as riding time, transfer time and waiting time, have been successfully derived from the data.

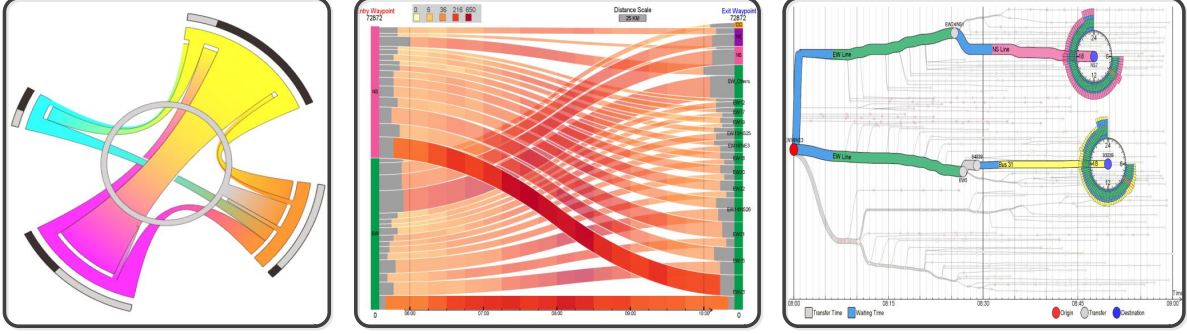


Figure 1.5: Novel visual representations proposed in this thesis from left to right: interchange circos diagram [151], waypoints-constrained OD view [152] and OD trip journey view [150], corresponding to the small-, medium- and large-scale OD-related analytical tasks, respectively.

• Visual Mapping

A family of novel visual representations that map the related information into more efficient visualizations have been proposed: Chapter 4 formulates a new model of circos figure, namely the interchange circos diagram, to present interchange patterns at junction nodes in a bundled fashion, and improve color assignments to respect the connections within and between junctions; Chapter 5 derives an unified visual representation called the waypoints-constrained OD view that integrates the flow map, heat map, and Sankey diagram to support the medium-scale OD-related analytical tasks; and Chapter 6 designs an integrated visual representation with there visualization modules, including isochrone map view, isotime flow map view and OD-pair journey view.

• View Transformation

Three visual analytics systems have been developed, allowing transport researchers to interactively explore and analyze the data: Chapter 4 develops an analytics system to help

transport researchers interactively study interchange patterns in a spatio-temporal manner, including 1) multi-spatial scales, i.e., from network junctions such as train stations to people flow across and between larger spatial areas, and 2) temporal changes of patterns from different times of the day; Chapter 5 proposes waypoints-constrained OD visual analytics, which allows users to interactively specify entry and exit waypoints in the transport network, and explore the corresponding OD patterns subject to the trajectories that successively pass through the waypoints; Chapter 6 also designs an analytics system that devises several interactive visual query methods, allowing transport researchers to easily explore the dynamics of PTS mobility over space and time.

Besides, each chapter presents case studies and interviews with transport experts to demonstrate the usability, effectiveness, as well as limitations of our visual analytics.

1.4 Thesis Organization

The remainder parts of this thesis present the contributions in more details: Chapter 2 summarizes the related work in the fields of movement pattern analysis, visual analytics for movement data and visualizing urban transport. Chapter 3 presents a detailed description of the input dataset and its applications in transport domain. Chapter 4 introduces a novel visual representation, namely *interchange circos diagram*, aiming at revealing commuter interchange patterns in a traffic network. Chapter 5 presents *waypoints-constrained OD view*, a novel unified visual representation designed to explore the OD patterns associated with the commuter trips successively passing through users-specified entry and exit waypoints in the transport network. Chapter 6 describes an integrated visual analytics with three visualization modules, i.e., *isochrone map view*, *isotime flow map view* and *OD-pair journey view*, aiming at presenting the mobility of a PTS. Chapter 7 concludes this thesis with a brief summary of the achievements made, and suggestions and thoughts for future work.

Chapter 2

Related Work

Urban public transport data can be considered a subset of movement data, which associates with people and goods flows in urban environments, and visualizing and exploring movement data have always been a hot research topic in many fields, including cartography, GIS, geography, transport, and visualization etc. This chapter mainly reviews the related work in the following three categories: 1) movement pattern analysis, 2) visual analytics for movement data and 3) visualizing urban transport.

2.1 Movement Pattern Analysis

Movement patterns generally describe any recognizable regularity or any interesting relationship within a group of moving objects in space and time [36], which are widely available in the nature world, e.g., a flock of birds, a herd of land animals, and a school of fish [100]. Reynolds [100] modeled the aggregated movements as interactions between the behaviors of individual moving object. However, this approach may not be efficient and robust anymore to handle the emerging vast amount of movement data, which can easily comprise thousands, and even millions of moving objects nowadays. Instead, various mining approaches have been proposed to effectively retrieve the movement patterns from raw data. In this section, we firstly

give a brief summarization of these mining approaches, and after that, we describe some examples that apply mined movement patterns to understand mobility-related phenomena.

2.1.1 Movement Pattern Mining

Data mining is a process of knowledge discovery that applies specific algorithms for extracting patterns from data [40]. Regarding movement data, various data mining techniques can be applied to discover usable knowledge about the behaviors of different types of the moving objects in both space and time [79]. Dodge et al. [36] summarized the mining techniques for movement data analysis, and suggested some potentially useful dimensions towards a taxonomy of describing and classifying movement patterns: generic vs behavioral, primitive vs compound, and group vs individual.

REMO [77], which analyzes the **RE**lative **MO**tion of individual object to the motions of all others by comparing their motion attributes over space and time, is a widely adapted in mining movement patterns. REMO firstly transfers the movement data into a matrix featuring the motion attributes, and then formalizes motion patterns by matching it with the formulated matrix. The authors illustrated how to extract simple movement patterns (e.g., constancy, concurrence and change) from movement data. They [80] further defined some more complex spatially-constrained movement patterns (e.g., flock, leadership, convergence and encounter), and discussed how to mine them with REMO. Two case studies were presented to demonstrate the effectiveness of REMO by identifying non-trivial and meaningful movement patterns in the movement data [78]. Later, Gudmundsson et al. [49] speeded up the mining process of REMO by devising new approximation algorithms derived from computational geometry methods.

Besides REMO, Giannotti et al. [47] formalized a general statement of movement patterns mining problem that depicts movement patterns as frequent behaviors in both time and space.

Based on the formalization, they proposed static and dynamic regions-of-interest based methods to extract frequent patterns from movement data. Pelekis et al. [95] introduced a set of distance operators to compute similarity between two trajectories, which can be employed not only to query similar movements, but also to cluster and classify the movements. Dodge et al. [35] proposed another method to segment and classify movements, which extracts local features from individual movement and compares them with global descriptors computed from an entire movements set.

2.1.2 Movement Pattern Applications

Movement patterns can help understand mobility-related phenomena, which is highly important in many domains. For instance, Brockmann et al. [25] assessed the circulation of bank notes in USA, and discovered that the features of human travel can be quantitatively accounted, where “*the distribution of traveling distances decays as a power law*”. The discovery of human movement pattern can be applied to study the spread of human infectious diseases.

In particular, when the movement patterns represent aggregated abstraction of many people trajectories within urban environments, they are highly useful means in the domains of city planning, transport design and traffic management. There has been an extensive amount of related works on analyzing the urban traffic data, thus here we give a brief summarization of them with an example of Microsoft’s urban computing, which aims at acquiring, integrating, and analyzing big and heterogeneous data generated by a diversity of sources in urban spaces to tackle the major issues that cities face [153]. T-dirve [148, 149, 147] is an example of urban computing applications, which is a smart driving direction service that finds the fastest route to a destination based on mined directions from historical taxi trajectories. And by associating the taxi trajectories with points of interests (POIs), they can further cluster the city into regions and estimate the region functionality [146].

2.2 Visual Analytics for Movement Data

There have been a long history in geography to develop interactive visual tools to explore and analyze movement data that can be mapped geographical space. Gahegan [43] outlined four barriers in the development of visualization tools for analyzing movements in geoscience: 1) rendering speed, 2) perceptual anomalies, 3) approaches and mappings, and 4) user orientations. Correspondingly, MacEachren and Kraak [87] suggested the following primary themes in geovisualization field: 1) visual representation of movements, 2) integration of visualization with computational methods 3) interface design and 4) cognitive/usability studies. These themes can also be applied when developing visual analytics tools for movement data.

Recently, visual analytics of movement data has also been a hot research topic in the visualization community, and extensive visual analytic tools have been developed to present the spatial and temporal semantics of movement data. Andrienko et al. [8, 11] presented a structured survey of the state-of-the-art visual analytics methods, tools and procedures. Here, we categorize the related works into temporal-oriented, spatial-oriented, spatio-temporal-oriented, and pattern-oriented visual analytics of movement data.

2.2.1 Temporal-Oriented Visual Analytics

Extensive amounts of efforts have been devoted to develop visualization methods to meet the needs of analyzing and understanding time-oriented data, i.e., the data is varying over temporal dimension, see [2, 3] for comprehensive surveys. In particular, Aigner et al. [2] highlighted that the time dimension could be linear or cyclical, so different visualization strategies could be adopted for different situations accordingly. For example, Havre et al. [56] proposed The-meRiver that depicts temporal changes in thematic strength from left to right, which is suitable to represent linear time data, e.g., the topic competition on social media over time [145]. On

the other hand, Weber et al. [136] proposed spiral graph that visualizes temporal changes in a circular manner, which better supports the identification of periodic structures in the cyclical time data, e.g., sunshine intensity variation over days.

Movement data is ubiquitously time-oriented exhibiting both linear and cyclical patterns. For instance, the positions of movements are changing with time in a linear manner, whereas the travel times between an origin-destination pair is cyclical over days. To present the movement changes in temporal space, various novel visual encodings have been proposed. For instance, Tominski et al. [122] proposed spiral-based icons that map temporal dependencies into a 3D dimension. Burch et al. [26] designed AOIRiver, which combines ThemeRiver and Sankey diagram techniques to visualize the dynamic eye gaze movements over time.

Besides these novel visual encodings, some traditional visualization techniques can also be applied, e.g., isochrone maps and cartograms.

Isochrone Maps. Isochrone maps are traditional visual representations used in transport and urban planning, which display areas of equal travel time from a starting location in certain time periods [137]. It usually employs contour lines/colors in its representation, and can be easily overlaid on geographical maps for depicting time-related information such as accessibility [155, 86].

Cartograms. While isochrone maps display reachable areas from a particular location, cartogram depicts proximities, such as time distance or travel costs, between locations in an Euclidean space [112, 115]. A comparison of a cartogram with a geographical map can show the distribution of proximities over space, while a comparison between cartograms drawn at different times can reveal changes over the period [113]. Many algorithms have been developed to transform geographical space into temporal space, including multi-dimensional scaling (MDS) [1] and timespace mapping [16]. Later, Shimizu and Inoue [113] showed that MDS and

network mapping algorithms can actually be formulated into a generalized time-space transformation solution.

2.2.2 Spatial-Oriented Visual Analytics

According to the First Law of Geography, where “*everything is related to everything else, but near things are more related than distant things*” [119], spatial-oriented analysis of movements can be formal quantitative studies of movements that manifest themselves in space, which should be focused on location, area, distance and interaction [13]. The main problem of spatial-oriented visual analytics of movement data is the visual clutter displayed on the screen due to the large-size and high-complexity properties of the data. To address the visual clutter problem and facilitate the analysis and exploration, appropriate aggregation and generalization of movement data is generally employed [8]. In general, following methods can be applied to summarize the movements in spatial domain.

Geographical Partition. Partition the geographical space into regions and then summarize the movements in-between regions can effectively reduce the cluttering problem. Typically, the geographical space can be partitioned according to the administrative units, e.g., states or cities. Besides, regions can also be directly formulated from the movement data. For instance, Guo [50, 53] developed a spatially graph-based partition method that can construct a hierarchy of geographical regions from the movement data. Besides, Andrienko and Andrienko [10] proposed a point-based partition method by extracting characteristic points from the movement data and then grouping the points into regions according to their spatial proximities.

Density Estimation. Besides geographical partition, density estimation is another frequently applied summarization method to address the cluttering problem. Based on kernel density estimation (KDE) [114], density maps can be generated as a mean to summarize large amount of trajectory paths in space, so that to overview the distribution of moving objects. Hurter

et al. [64, 63] developed FromDaDy, a multidimensional KDE visualization tool providing a brush/pick/drop paradigm for users to explore large amount of aircraft trajectories across multiple views. Willems et al. [141] also developed a density-map-based interface to visualize vessel movements with large kernels to overview spatial utilization and reveal vessel highways, and small kernels to show speed variations of individual vessel. They further improved the system with interactive functionalities, such as specifying density fields, filtering the trajectories, and exploring customized versatility of the movement data [106, 105, 107, 140].

Clustering. What's more, clustering movements into groups and visualizing the groups of movements can also address the cluttering problem by reducing the morphological complexity, while accentuate important movement patterns. A variety of clustering methods have been proposed to group the movements, including progressive clustering [102], traffic-oriented and trajectory-oriented clustering [6], classifier-based clustering [9], vector-field based clustering [41], and traditional hierarchical clustering [154]. Vrotsou et al. [130] compared geometry-based, density-based, and property-based clustering methods, and suggested different approach should be employed based on the analysis tasks.

2.2.3 Spatio-Temporal-Oriented Visual Analytics

The entanglement of movements, i.e. they are happening in both space and time, adds to the difficulty of analyzing and visualizing the movements [97]. Spatio-temporal-oriented visual analytics of movements attempt at presenting the moving objects over time as they move through space, and understanding the correlations between space and time. These analytics generally study three basic kinds of questions in analyzing the movement data: (1) when + where $\rightarrow$ what, (2) when + what $\rightarrow$ where, and (3) where + what $\rightarrow$ when [96].

3D View. To visualize both the spatial and temporal attributes, 3D views with the height axis to denote time are frequently employed, which are normally refereed as space-time cubes

(STC) [55]. Kwan [75] showed that 3D views can simultaneously present the spatio-temporal dimensions of movement patterns, and facilitate the spatial-relation identification and different subgroups patterns comparison. Among various STC applications, GeoTime [69] is perhaps the best known software, which can be employed to display and track events, objects and activities with combined spatial and temporal information (Figure 2.1(a)). Besides, Tominski et al. [123] proposed to stack up 3D trajectory bands to visualize corresponding attributes of the movement data (Figure 2.1(b)).

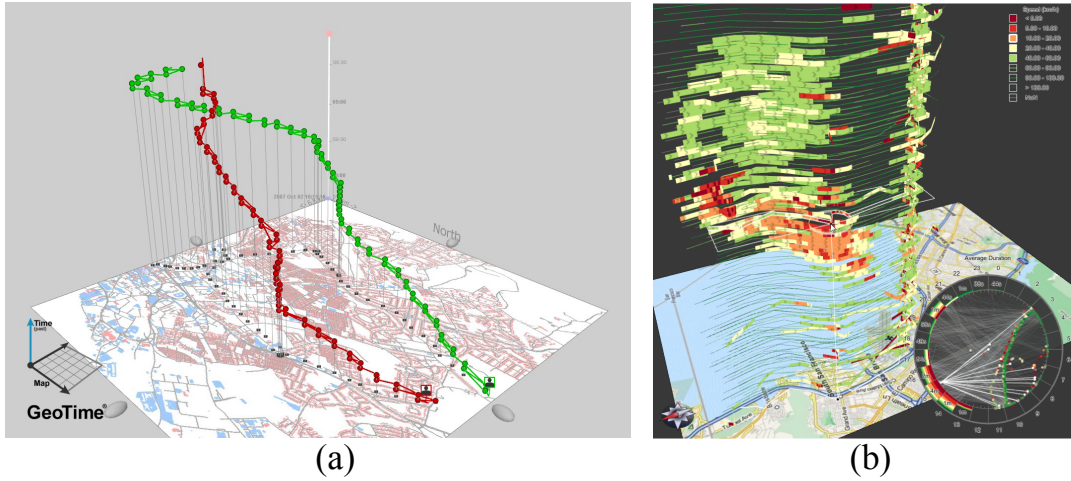


Figure 2.1: Examples of 3D visualizations of movement data: (a) GeoTime [69] and (b) stacking-based visualization [123].

Animations and Small Multiples. However, 3D views are normally considered deficient in visual analytics field possibly due to occlusion, perspective distortion, increased complexity of navigating and bad text legibility problems [29, 48]. Thus, animations [19] or small multiples [124] of 2D views are also frequently employed to visualize the temporal change of movements over a period. For instance, Guo et al. [51] developed a visualization system with a small multiple display to present both the spatial and temporal patterns in the movement data. However, the animations and small multiples have evident drawbacks: with animations, temporal variations are hard to track and identify [125]; and with small multiples, details are difficult to observe due to the smaller display size [23].

2D and 3D Comparison. Many experiments have been carried out to compare 2D and 3D visualizations of movement data. Kristensson et al. [72] teased out that comparing to 2D, 3D visualization results in higher error rates for simple and direct queries, while performs better when observing complex spatio-temporal patterns. Amini et al. [5] further showed that interactions have significant effects on 2D and 3D visualizations: “scrubbing” the timeline for 2D, while camera navigation for 3D.

2.2.4 Pattern-Oriented Visual Analytics

Various data mining techniques have been proposed to extract movement patterns in the data, see Section 2.1.1 for more details. Pattern-oriented visual analytics present these extracted movement patterns to users for interpretation, evaluation and synthesis [8]. The visual analytics can allow users to catch sights of noteworthy knowledge by directly looking and interacting with the visualization. Here we demonstrate how pattern-oriented visual analytics work with following examples in visualizing OD patterns.

Visualizing OD Patterns. OD pattern summarizes the movements between locations, which could be mostly common among various movement patterns. A family of visualization techniques can be employed to visualize OD patterns in movement data, including:

- *Flow Map.* Flow map joins origins and destinations by straight/curved arrows with line width indicating aggregated flow volume [120, 121], which maybe the most common visualization approaches for presenting OD data. Other than joining the ODs directly, Phan et al. [98] proposed flow map layout method that automatically clusters nodes into a tree-like hierarchical structure, and then bundles neighboring flow lines to present the general flow trend. Verbeek et al. [127] further improved the layout by computing crossing-free flows based on a spiral-tree method.

However, flow map has certain issues: 1) visual clutter would easily occur when applied to large amount of movement data; 2) longer flow lines can easily overlap and occlude shorter flow lines [143]; 3) modifiable area unit may generate different aggregation (or even wrong) patterns [45, 92]; and 4) the aggregations are often dramatically different in size [52]. To address these issues, partition the geographical space into larger regions before applying the flow map may effectively resolve the visual clutter problem [50, 10].

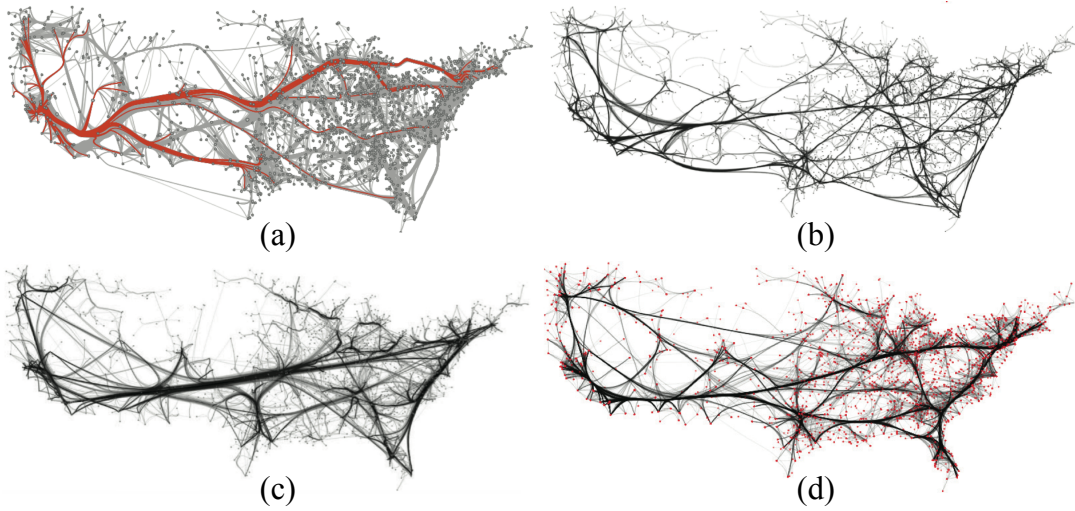


Figure 2.2: Examples of edge bundling techniques applied to visualize US migrations: (a) geometry-based [34], (b) winding roads [76], (c) force-directed [60], and (d) kernel density estimation-based [62].

- *Graph Visualization.* Another approach to visualize OD data is to employ graph visualization techniques, which denote origins and destinations as nodes, and thus can represent the flows in-between origins-destinations with directed edges. For instance, extended arc diagrams were proposed to visualize the traffic flows between different stops in the public transit network [90]. For large OD pairs, the graph visualization may also suffer from cluttering problem with excessive edge crossings. Edge bundling can effectively facilitate the visualization by grouping edges into bundles, and a family of bundling methods have been proposed, including geometry-based [34] (Figure 2.2(a)), winding roads [76] (Figure 2.2(b)), force-directed [60] (Figure 2.2(c)), skeleton-based [39] and

kernel density estimation-based [62] (Figure 2.2(d)).

- *Matrix Visualization.* Mathematically, OD data is an M -by- N matrix, summarizing flow volumes from M origins to N destinations. Thus, another commonly employed approach to visualize OD data is matrix visualization, which can produce better readable representations than the graph-based approach when the OD pairs are large [46]. Guo et al. [51] showed that the rows and columns of OD matrix can be sorted and re-ordered to reveal the apparent clusters. And Wood et al. [142, 143] improved the OD matrix visualization by dividing the geographical domain into regular grids to preserve the spatial structure of the origins and destinations.
- Boyandin et al. [22, 21] specifically focused on visualizing and exploring temporal changes in the OD data. They designed Flowstrates [22], which shows origins and destinations on two separate geographical maps on the left and right, respectively, and presents temporal changes by a heat map in the middle.

2.3 Visualizing Urban Transport

Thanks to the advancement in sensing devices, such as mobile phones, GPS devices, laser scanners and RFID cards, collecting urban transport data have become easier. Many visualization works aiming at improving the understanding of human mobility have been developed with state-of-the-art information technologies in recent years. More specifically, these visualizations can be classified into three categories: transport network visualization, urban traffic simulation, and urban traffic visual analytics.

2.3.1 Transport Network Visualization

Urban transport networks have become increasingly complex. Only considering the public transport networks in metropolitans like London and New York, they can easily comprise hun-

dreds of subway stations and thousands of bus stops. The complexity brings challenges to develop visualizations to explore and analyze the network.

In 1931, Mr. Harry Beck created the famous London Underground Tube Map that connects the subway stations with straight and 45 degree lines [44]. Since then, the idea has inspired the designing of most metro maps around the world [94], and they are generically referred as schematic maps. Many algorithms have been proposed to automatically generate appealing schematic maps, including mixed-integer programming [91], focus+context [133], spatially efficient [144] and transit-centric [32].

Schematic maps basically overview the transport network's essential topological relations across stops. The maps serve as visual aids to assist commuters navigate the network and plan routes from their origins to destinations [89]. Through a user study, Bartram [18] confirmed that commuters can navigate the transport network efficiently and accurately with schematic maps. Meilinger et al. [88] further explored and confirmed the value of schematic maps for easier wayfinding and self localization.

2.3.2 Urban Traffic Simulation

To enhance the sense of immersion, simulation techniques are often employed to visualize the dynamic motions of vehicles over time in 3D virtual cities. Two steps are often carried out to construct a realistic simulation of traffic flows.

The first step is to construct an accurate road map that constrains the vehicle motions. Wilkie et al. [138] proposed an efficient approach to generate a geometrically and topologically consistent 3D road map from massive GIS data, where important road features are identified, including highways, legal merge zones, and intersections. Similarly, Shen and Jin [111] also simulated the traffic flows based on an urban arterial network with detailed information like signalized crossing, merging and weaving areas.

After synthesizing the road map, modeling and simulating the traffic can be implemented. Helbing [58] broadly classified traffic simulations into two categories: macroscopic models, e.g. [110], and microscopic models, e.g. [108, 111, 139, 30]. The main advantage of the macroscopic models is the computational efficiency, while microscopic can model the micro-level behaviors of the vehicles, like lane-changing and vehicle-acceleration. Sewall et al. [109] generated interactive visual simulation of large-scale traffic flows by coupling the macroscopic and microscopic models together. In order to integrate these two models, they proposed averaging and Poisson-process techniques to make smooth transitions.

2.3.3 Urban Traffic Visual Analytics

Urban traffic visual analytics employ state-of-the-art visualization techniques to investigate and analyze traffic patterns from urban traffic data in cities. It is a cross-disciplinary research field that integrates knowledges in both transport and visualization domains. A family of studies have shown that visual analytics can not only facilitate the understanding of urban dynamics and human activities, but also enhance traffic management and assessment.

Among these studies, Guo et al. [54] designed TripVista to investigate and analyze microscopic traffic patterns and abnormal behaviors at road intersections from spatial, temporal and multi-dimensional perspectives. Liu et al. [83] proposed novel visual encoding schemes to display, compare, and evaluate route diversity in-between given OD pairs in real taxi drivers trajectory data, which can be applied to not only suggest better routes, but also analyze traffic bottlenecks. And Liu et al. [85] developed VAIT, a visual analytics system for intelligent transport that supports visualization and analytical queries of large traffic data.

More recently, Chu et al. [31] visually explored hidden themes of taxi movements in a city. Wang et al. [134] designed visual analytics to support the analysis of traffic congestions in city scale (Figure 2.3). They [135] further studied traffic congestion's correlation with traffic

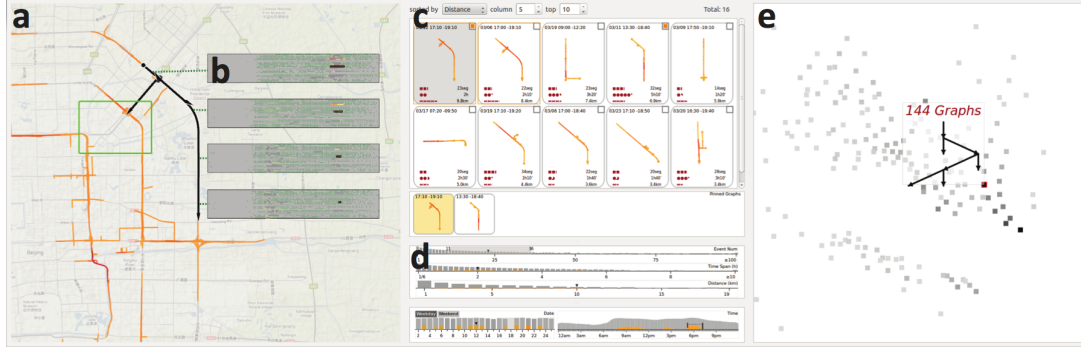


Figure 2.3: An overview of the visual analysis system to explore traffic jam based on urban traffic data proposed by Wang et al. [135].

flows on neighboring links. Meanwhile, Ferreira et al. [42] developed new models to support interactive spatio-temporal queries of events from large traffic data set. Doraiswamy et al. [37] proposed a visual exploration interface that can automatically identify interesting events and trends by computing the topologies in the urban traffic data (Figure 2.4). Besides, Wang et al. [131] designed an visual interface to evaluate traffic situations on particular roads, which was supported by a bi-directional hash structure.

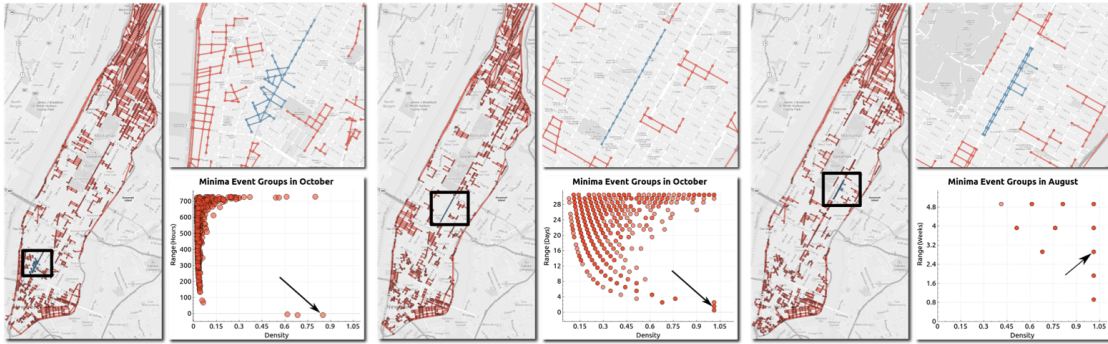


Figure 2.4: Minima events identification in NYC using topological analysis in urban traffic data proposed by Doraiswamy et al. [37].

2.4 Inadequacy of Existing Methods

Though so many related works exist, they are not focusing on the multi-scale OD patterns that this thesis aims to resolve. In particular, the inadequacy of these existing methods can be

summarized as follows:

- **Usability and applicability:** Most existing movement pattern analysis works are not directly and practically usable and applicable as visual analytics this work aims to achieve, since visual analytics are aimed to facilitate domain experts in exploring and analyzing the data, and to inform the general public with useful traffic and activity information, while the knowledge retrieved from movement pattern analysis still needs to be presented to end users in more intuitive and effective ways.
- **Data source:** Most existing urban traffic data visual analytics systems are developed to study GPS-based movement data, while this thesis is focusing on urban public transportation data. The differences between these two data sources include: 1) the public transportation is multi-modes, such as bus, subway and walk, while the GPS movement data normally records only single travel mode; 2) the public transportation data is more sparse, since the tapping events are only recorded at stops, while the GPS movement data updates frequently in every 30 seconds (or 1 minute); 3) many observations in the public transportation correspond to one vehicle movement, while every record in the GPS movement data is normally one vehicle. A more detailed description of properties of the public transportation data is summarized in Section 3.3. In summary, the existing urban traffic data visual analytics systems cannot be directly applied in this work.
- **Multi-disciplinary:** This work is a collaborative work with transportation researchers, since the project is mainly funded by research grant that supports also a number of the transportation researchers, whose offices are actually co-located together with that of the author. Hence, the author has close discussions from the very beginning. In comparison, the other visual analytics systems are mainly led by computer scientists and visualization specialists that mostly lack of domain knowledge in transportation field.

Chapter 3

Massive Urban Public Transport Data

The process of global urbanization calls for an increasing demand of mobility in cities, and more researchers and planners are realizing the importance of developing urban PTSs, which can provide shared and massive transport services that are essential for general public. As such, many researches have been carried out to study and analyze the underlying movement patterns of commuters utilizing PTSs. Thanks to recent development of ubiquitous sensing devices, massive amounts of urban public transport data are collected in an automatic and pervasive way, providing transport researchers a preferable choice to study movement patterns.

This chapter firstly introduces relevant terminologies from transport research. After that, we present an overview of the urban public transport dataset employed in this thesis, followed by a description of the data characteristics and applications in transport domain.

3.1 Transport Terminologies

In the following, we list down some common terminologies about PTS [93] to facilitate the discussion:

- A *transport network* consists of roads and subways, and is usually modeled as a directed

graph data structure, where nodes are stops (metro station platforms and bus stops) with geographical locations, and directed edges connect neighboring nodes;

- A *transit route* is a sequence of nodes and edges, starting and ending at bus/subway terminals;
- A *transit line* is a public transport service offered by a certain transport mode, e.g., a bus line and a subway line. There are two kinds of transit lines: *bidirectional* with two transit routes between two distinct terminals, and *cyclical* with one single transit route starting and ending at the same terminal;
- A *trip* refers to an individual transit route service taken between two stops/terminals;
- A *transfer* refers to a change of transit route services; it could happen at the same location (e.g., a bus stop), or between two different but neighboring locations (e.g., between different subway platforms or from subway to bus); and
- A *journey* is a commuter travel from an origin to a destination in the PTS; it could comprise multiple trips and transfers.

3.2 Input Public Transport Dataset

The work described in this thesis is based on the following input dataset:

3.2.1 EZLink Data

The EZLink data records commuter trips in Singapore PTS over one week in 2011. The PTS includes a metro system referred as mass rapid transit (MRT) and a public bus system, where commuters use their EZLink cards to tap on card readers on buses or entries/exits at MRT stations to go in/out of the PTS. The card reader system records every tap event, and also

considers transfers between bus and MRT services: if the transfer time is ≤ 30 minutes, the two trips will be sequenced together and identified by consecutive transfer number.

The EZLink data has about five million trip records in total for one day. For each trip, following information of the trip is recorded:

Column 1: **Card ID**, an anonymous ID of the EZLink card.

Column 2: **Journey ID**, ID of the commuter journey.

Column 3: **Commuter Type**, type of the commuter, among *Child/Student*, *Adult* and *Elder*.

Column 4: **Transport Type**, type of the transport mode, either *BUS* or *RTS* (Rapid Transit System, including MRT).

Column 5: **Service Number**, transit route number for a BUS trip, such as *179* and *199*; *NULL* for RTS trips.

Column 6: **Direction**, direction of the BUS route; *NULL* for RTS trips.

Column 7: **Bus Registration Number**, unique number registered for a bus in the BUS route, e.g., *9088* for a bus in transit route *179*; *NULL* for RTS trips.

Column 8: **Tap-In Stop**, stop ID for the tap-in stop.

Column 9: **Tap-Out Stop**, stop ID for the tap-out stop.

Column 10: **Ride Date**, date of the trip starts, e.g., *11/04/11*.

Column 11: **Ride Starting Time**, time recorded when the commuter enters the public transport system, i.e., boarding a bus or entering an entrance of a MRT station.

Column 12: **Ride distance**, trip distance measured in *kilometer*.

Column 13: **Ride Time**, time spent in the PTS of the trip in *minute*, i.e., tap-out time subtract tap-in time.

Column 14: **Fare Paid**, fare paid for the trip in *Singapore dollar*.

Column 15: **Transfer Number**, the transfer sequence number of a journey, e.g., from Changi Airport to NTU, commuters need to take XX line, then switch to route 179 at Boon Lay. Thus, the XX trip will be assigned a transfer number 0, and 179 trip transfer number will be assigned 1.

Figure 3.1 shows an example of EZLink records revealing a child/student journey traveling to NTU (stop 27321) in the afternoon. The journey consists of three consecutive trips: First, a bus trip taking bus route 98M from stop 28329 to stop 28319; second, a RTS trip from MRT station A to MRT station B; and third, a bus trip taking bus route 179 from stop 22521 to stop 27321. Besides reconstructing the travel paths, one can also extract other information from the records, such as travel times and costs.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
102456972513	2000088401530800	Child/Student	Bus	98M	1	9075	28329	28319	11/4/11	15:53:17	0.6	6.75	0.36	0
102456972513	2000088401530800	Child/Student	RTS	?	?	?	STN Jurong East	STN Pioneer	11/4/11	16:03:28	5.6	12.85	0.15	1
102456972513	2000088401530800	Child/Student	Bus	179	1	9088	22521	27321	11/4/11	16:25:29	1.1	3.883	0.07	2

Figure 3.1: An example of the EZLink records revealing a child/student journey traveling to NTU (stop 27321) in the afternoon.

3.2.2 Public Transport Network

The network includes the road and MRT network in Singapore, which can be modeled as a directed graph that contains around 47.9k nodes and 79.8k edges in 2011. In detail, the MRT network contains four transit lines at that time, i.e., the XX line with green color, the YY line with red color, the ZZ line with purple color and the WW line with yellow color as shown in Figure 3.2, respectively.

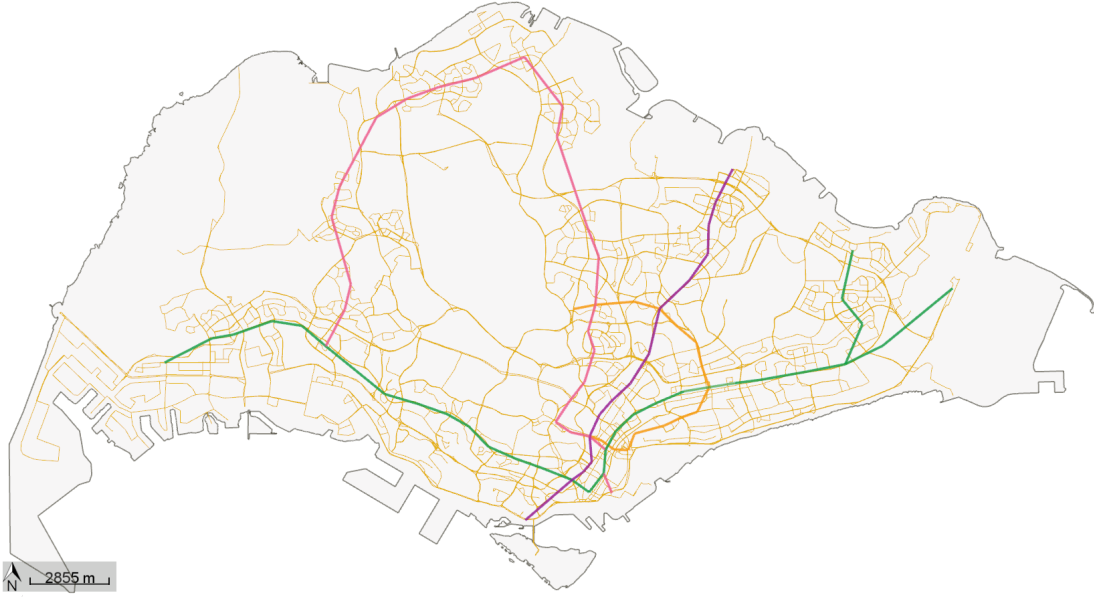


Figure 3.2: Singapore public transport network in 2011. Thin brown lines indicate the bus network, while thick green, red, purple, and yellow lines indicate the XX, YY, ZZ and WW MRT lines, respectively.

3.2.3 Transit Schedule

The transit schedule data includes the transit routes, stop facilities and transit schedule information in 2011. Besides the four MRT lines, the transit routes data also includes around 350 bus routes in operation. There are in total around 4.8k stop facilities, including both bus stops and MRT stations, each with the attributes of a geographical position, a name, a reference ID, and a related edge connection in the PTS network. The transit schedule information is basically a timetable describing the scheduled times of each bus/train leaves its starting terminal, and reaches each stop along its transit route.

3.3 Data Characteristics

The input EZLink card data consists of a massive collection of commuter trajectories traveling in the urban public transport network, where each trajectory can be represented as an ordered

sequence of observations on individual commuter [61]. The data can be considered as a special case of movement data that describes a set of moving objects whose positions or geometric attributes change over time [36].

As such, the input urban public transport data exhibits some general properties of common movement data, such as:

- **Big Data Volume:** The data consists of about four million of commuter trips each day, and there are 7 days records in total. The raw data came in CSV files with a total size of approximately 4.2 GB. Notice that the public transport data only takes a single line to record one trip. Comparing with other movement data that records many points along the path for each trip, the file size is much smaller, but the information is equally plentiful.
- **Complex Network:** The public transport network includes thousands of nodes and edges. Besides, the input dataset also includes around hundred MRT stations and thousands of bus stops. These nodes, edges and stops can be integrated and modeled as a directed complex network.
- **Integral and Organic:** From the input dataset, one can estimate both independent information on each commuter, such as movement speed and direction; meanwhile, one can also derive movement patterns on the entire data, such as density distribution [12].
- **Spatial Variation:** Variation of movement attributes, such as speed and density, exists in the spatial dimension. Figure 3.3 presents a map depicting the travel efficiencies of all bus routes in the period of 08:00 - 08:15. Here the travel efficiencies are quantized and color-coded in five ranges: below -1.5σ , $[-1.5\sigma, -0.5\sigma]$, $[-0.5\sigma, +0.5\sigma]$, $[+0.5\sigma, +1.5\sigma]$, and above $+1.5\sigma$. Low travel efficiencies can be observed at downtowns, while high efficiencies are observed at highways and suburban areas.
- **Temporal Variation:** The variation also exists in the temporal dimension. Figure 3.4 shows the distribution of alighting commuter volume at XXX MRT station during two

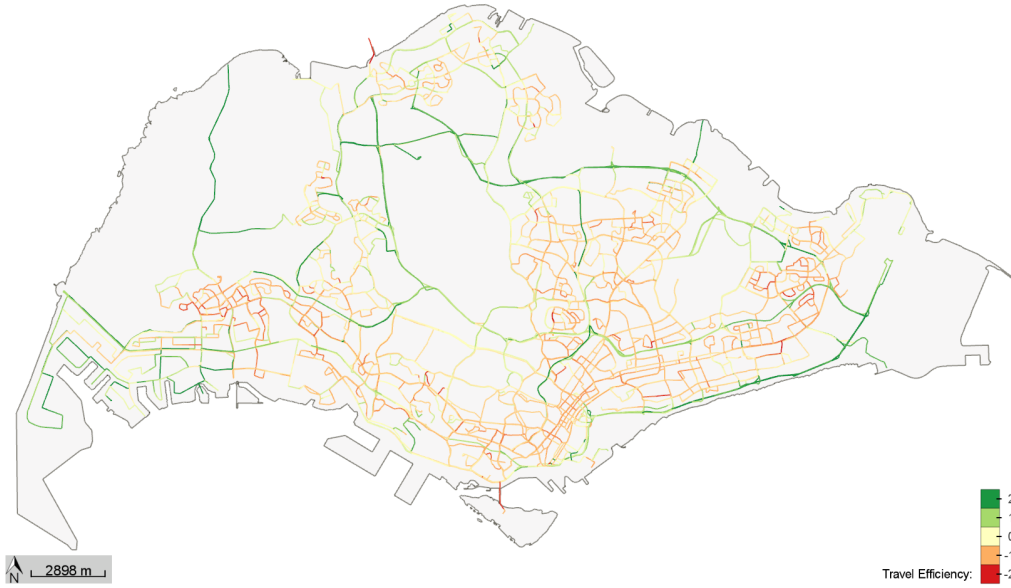


Figure 3.3: Road network travel efficiency in the period of 08:00 - 08:15. Green and red colors indicate high and low travel efficiency w.r.t. mean, respectively.

different periods, where left is in the period of 08:00 - 09:00, and right is in the period of 11:00 - 12:00. By comparing the two views, one can clearly find that more frequent and larger volume of commuters are alighting in the period of 08:00 - 09:00.

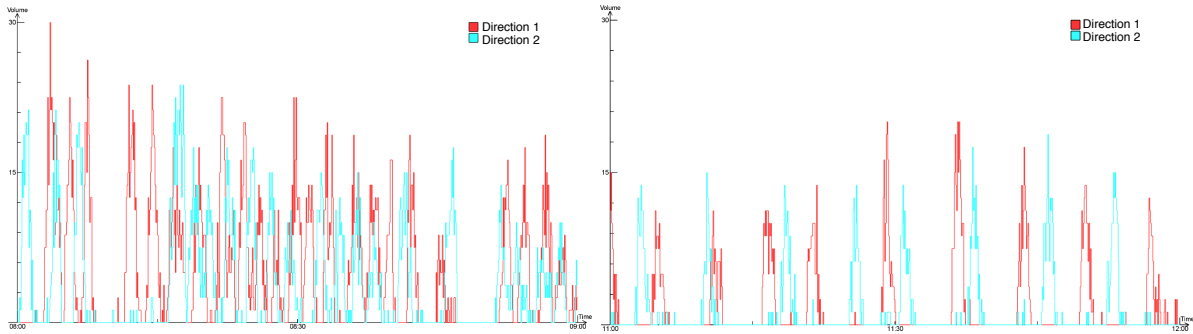


Figure 3.4: Temporal variation of alighting commuter volume distribution at XXX MRT station during different periods. Left: 08:00 - 09:00; Right: 11:00 - 12:00.

On the other hand, comparing with other kinds of movement data, the input public transport data owns some unique characteristics including but not limited to:

- **Sparsity:** In terms of space, the tapping events are only recorded at stops; in terms of time, the tapping events only happen when the vehicles approach/arrive stops. Thus, the

EZLink data is sparse in both space and time dimensions. On contrast, many other kinds of movement data are recorded in a quasi-continuous manner, such as periodically (e.g., in one-minute frequency) updated taxi data.

- **Localization:** Commuters taking PTS are aligned with vehicles, i.e., they travel on buses or MRT trains. Thus, the trip paths are only limited to the public transit services routes, instead of the whole transport network. On contrast, most other kinds of movements, such as taxi trajectories, can move through the entire transport network.
- **Uncertainty:** For each trip, only the starting and ending locations, and the corresponding times are observed. Although the speeds and locations of vehicles in-between two consecutive stops can be probed from the data, they are uncertain.
- **Multi-modes:** Although commuters are only involved in PTS, they can travel through the public transport in multiple modes such as bus and MRT, and they can also switch between these modes through walking.

3.4 Data Applications

With the plentiful information in the records, the EZLink data can provide transport researchers insights and knowledge of how commuters utilize the PTS. In the following, we discuss how the input dataset can be applied in each step of the classical four-step model of transport planning and forecasting:

Step 1: Trip Generation

In the trip generation step, transport researchers predict the number of trips originating in or destined for each traffic analysis zone. Many factors may affect the estimation in the zone, such as land uses, household demographics and social-economical factors.

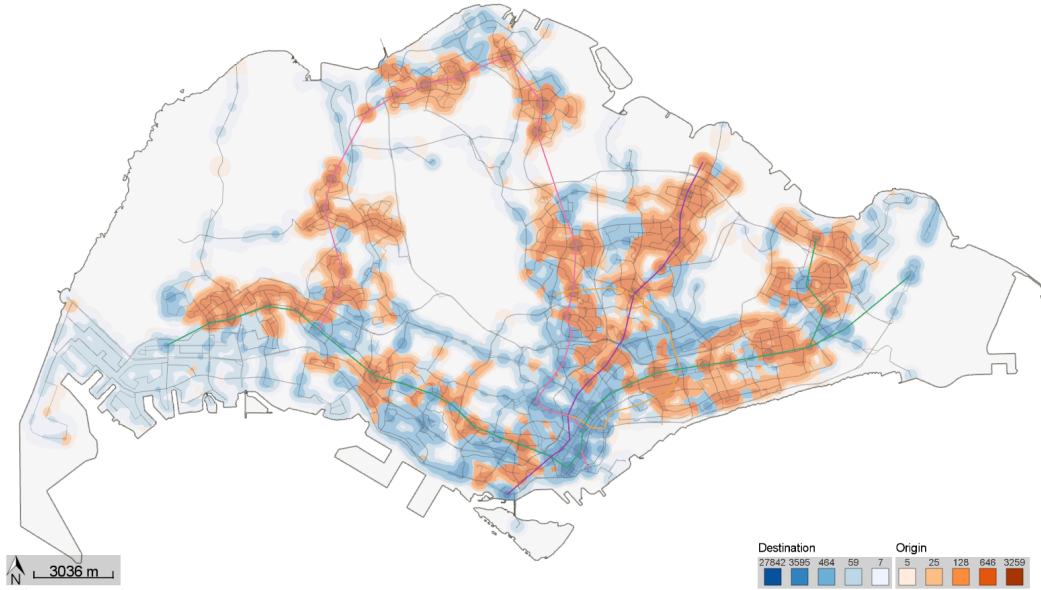


Figure 3.5: Origin and destination density in the morning retrieved from the input dataset, with red color indicates more commuters originating in and blue color indicates more commuters destined, respectively.

However, generally the researchers estimate the number based on only a subset of these factors, and thus the results are normally not optimal.

As the EZLink data contains the commuter boarding and alighting location information, one can accurately compute the number of trips in each zone. For example, Figure 3.5 presents a composited density map showing commuter origins and destinations in the morning. More commuters starting from residential areas can be observed, while more commuters are ending at downtowns and industrial areas.

Step 2: Trip Distribution

After generating the number of trips in each zone, transport researchers match origins to destinations in the trip distribution step. This is generally done with a gravity model function, where the number of trips between two zones is negatively associated with travel cost (including distance, time, and money) between them, and positively associated with the amount of activities in each zone [129]. EZLink data can also provide the distribution information as the origin and destination are matched in each record.

Step 3: Mode Choice

In-between a particular OD pair, commuters may have multiple transport modes to choose, including private transport like car, and public transport like subway and bus. In the mode choice step, the proportion of each particular transport mode in-between an OD pair is estimated. EZLink data also contains the transport mode information, specifically referred as *RTS* or *BUS*.

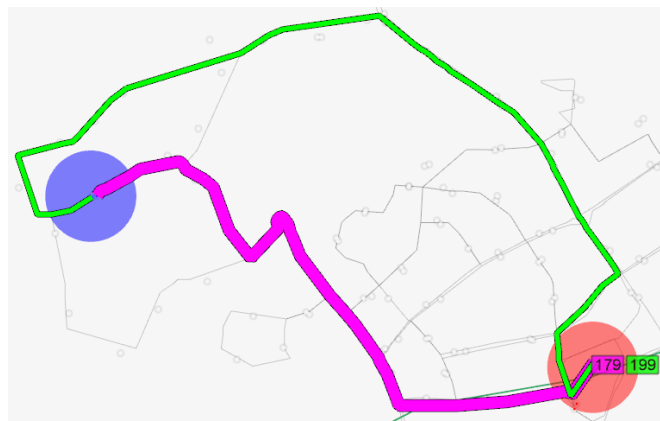


Figure 3.6: Multiple routes retrieved in-between red circle as the origin and blue circle as the destination from the input dataset.

Step 4: Route Assignment

Commuters may also have multiple routes to choose in-between an OD pair. In the route assignment step, transport researchers assign travel route to each of the trip. One can also assign the travel routes to each trip with information extracted from the EZLink data: For *BUS* trips, one can retrieve the bus number and then map the number onto the road network by referring to the transit schedule data; And for *RTS* trips, one can retrieve the boarding and alighting stops and look for shortest-path/shortest-travel-time travel path in the MRT network. Figure 3.6 illustrates an example of multiple routes retrieved in-between red circle as the origin and blue circle as the destination from the input dataset. Clearly we can observe that commuters choose either bus route 179 (pink) or 199 (green).

Chapter 4

Visual Analytics for Interchange Patterns

In this chapter, we focus on a novel aspect of visualizing and analyzing massive urban public transport data, i.e., the interchange pattern, aiming at revealing commuter redistribution in a traffic network. We first formulate a new model of circos figure, namely the *interchange circos diagram*, to present interchange patterns at a junction node in a bundled fashion, and assign the colors to respect the connections within and between junction nodes. Based on this, we develop a family of visual analytic techniques to help transport researchers interactively study interchange patterns in a spatio-temporal manner: 1) multi-spatial scales: from network junctions such as train stations to people flow across and between larger spatial areas; and 2) temporal changes of patterns from different times of the day. Our techniques have been applied to the input transport data, and we present also two case studies on how transport experts worked with our interface.

4.1 Introduction

A number of advanced data acquisition technologies have been developed recently for capturing movement data: location-positioning by cell phones and GPS, personalized user-tagged cards for public transport, and video analysis for people and vehicle flows. These technologies

benefit many scientific research disciplines, for example, the reconstruction of traffic flows from traffic sensors [108], and functional road models for traffic simulation [138]. However, such advancement also increases the data set size, thus making the problem of visualizing and exploring movement data to be nontrivial. Traditional methods [75, 69], which directly plot the object trajectories in 2D/3D, could simply fail because of visual cluttering and occlusion.

To address these issues, there are two major visualization approaches [8]: 1) *pattern extraction*, which applies knowledge discovery methods [80, 47] to find out motion patterns; and 2) *data aggregation*, which groups locations into regions and summarizes the movement data in a regional basis [50, 10]. Here we consider both strategies. In particular, we are interested in studying and visualizing a high-level aggregated motion pattern:

Interchange pattern, which describes how moving objects redistribute when entering and passing through a junction node in a traffic network.

Our formulation also considers the study of interchange patterns at different scales: train stations in a metro system, crossroads in a road network, or regional zones in a city.

Interchange pattern is a highly valuable means not only for unveiling mobility patterns, but also for assisting transport planning. For instance, interchange information can help reveal the road junction utilization and suggest crossroad redesign, e.g., adding a fork. A similar situation is also shared by the case of train stations, where interchange patterns can help improve the interior design of routes and platform connections within a station. At city scale, interchange patterns of people flow can help indicate longer distance trips or detours that are undertaken by some people, thus suggesting the transport efficiency for enhancing the road network design.

To support efficient visualization of interchange patterns that emerged from massive movement data, we propose a novel visual representation, namely the *interchange circos diagram*, for presenting the redistribution of people at junction nodes. This visual design is adapted from the circos figure [74], which was invented for examining the mutual relationships between

genomes. Incorporated with various advices from domain experts, we revise and customize the circos figure for presenting commuter interchange: a flyover ring to denote the junction node itself, bi-directional bundling to reduce visual cluttering, and an improved color assignment on linkages to enhance the visual connections between neighboring interchange circos diagrams. Our visualization techniques have been applied to real world movement data consisting of hundred thousands of trajectories, and two case studies on how transport experts applied our method are also presented.

4.2 Overview

This section first presents a formal definition on interchange patterns, and then overviews our system workflow.

4.2.1 Formal Definition: Interchange

An interchange pattern at a junction basically describes how moving objects redistribute when they go through the junction. Given a traffic network modeled as an undirected graph, say $G = (V, E)$, where V is the set of (junction) nodes in G and E the set of edges connecting neighboring nodes in V . When a moving object passes through a junction node, say $v \in V$, whose valency is n , it has $n + 1$ possible ways of entering the node. This is because it may come from v 's n connecting links, or from the dominion of junction v itself; these are the possible sources. Likewise, there are also $n + 1$ possible ways (sinks) of leaving junction v .

Hence, given the trajectory data, we first can identify a subset of trajectories that go through each node in V . Then, we can determine the incoming and outgoing links of each trajectory across a node, and summarize the interchange information at the node as a $(n + 1)$ -by- $(n + 1)$ matrix, which counts all the possible routes of going through the node.

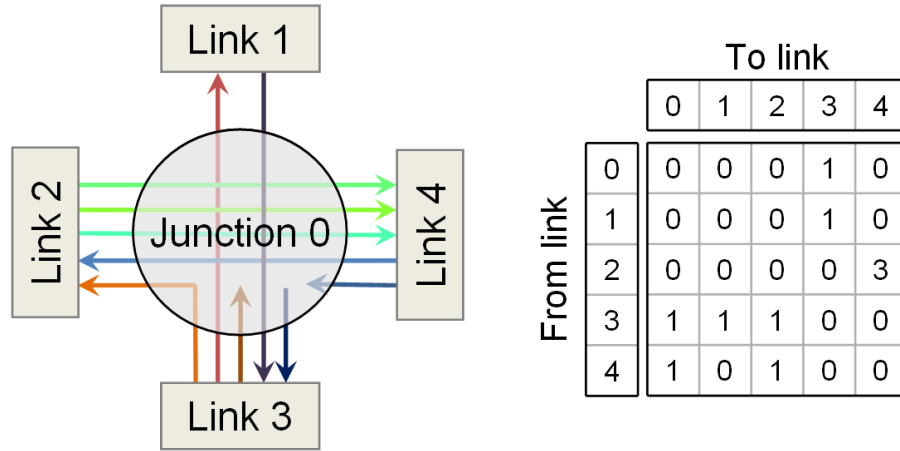


Figure 4.1: The interchange information (ten trajectories) at this junction can be summarized as a 5-by-5 matrix.

Figure 4.1 shows an example of a junction node with four links and ten trajectories. We can summarize its interchange statistics as a 5-by-5 interchange matrix. Note that the diagonal elements in the matrix are all zeros because we assume that no trajectories revert back to the same link in reality.

4.2.2 System Workflow

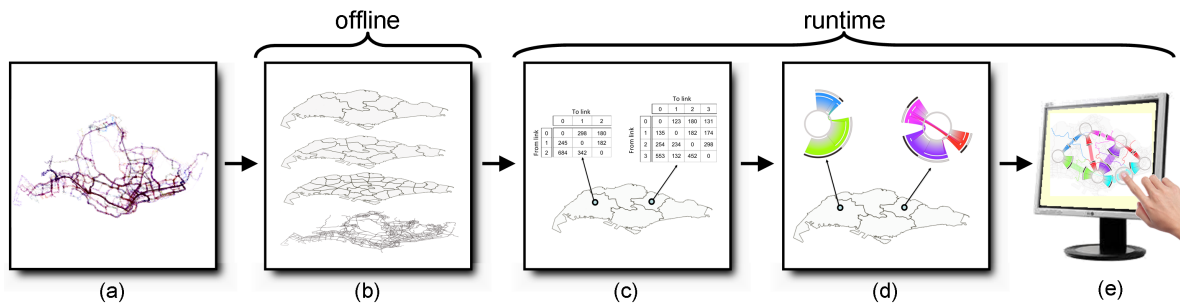


Figure 4.2: System workflow: from (a) a set of raw trajectory paths, to (b) traffic networks of different spatial scales, (c) interchange statistics, (d) interchange circos diagrams per junction node, and (e) our interchange visualization with user interaction.

Our system workflow consists of the following computational steps, see also Figure 4.2:

- Starting from the raw trajectory paths (Figure 4.2(a)), we first build a traffic network in the form of a undirected graph. It can be a road-level network (Figure 4.2(b)(bottom)), a

city-scale network (Figure 4.2(b)(top)), or a series of region-scale networks in-between. As for the finest-scale network, we can reconstruct it by examining the raw trajectory paths, while for the coarser networks, we can either reconstruct them by hierarchical clustering, such as those in [50, 10], or obtain the network structure directly from the domain experts.

- Then, for each traffic network, we determine per link (between pairs of neighboring junction nodes) two sets of trajectories (per movement direction along the link) that go through the link. Next, we partition the total time period covered by the the trajectories into equal time intervals, say 15 minutes, and precompute an interchange matrix for each time interval per junction node.
- After the user interactively chooses a period of time over a day, our system can retrieve and sum up the interchange matrices corresponding to the related time intervals that made up that time period. By this, we can quickly produce summarized interchange matrix (Figure 4.2(c)) at any junction upon user request. After that, an interchange circos diagram is constructed from the matrix and presented in the visualization (Figure 4.2(d)), see Section 4.3.
- Lastly, our interface supports also a family of visualization and user interaction techniques to explore various aspects of the interchange patterns, see Section 5.4.

4.3 Interchange Circos Diagram

Transport domain experts expect the following information when examining interchange patterns: (1) absolute and relative flow volumes across different pairs of links at a junction, (2) ratio of total incoming and outgoing flow volumes of each link, (3) flows starting/ending at the junction itself, (4) flow directions, (5) correspondence to the geographical nature of the data,

and (6) temporal and spatial variations of the interchange patterns. Hereby, we design a novel visual representation to capture these features.

This section first presents the idea of the original circos figure, and then develops it into the interchange circos diagram to present interchange patterns. Then, we present how the interchange circos diagram is implemented, and compares it against existing visual representation.

4.3.1 Circos Figures

The circos figure was invented by Krzywinski et al. [74] for examining the mutual relationship among genomes. After constructing a two-dimensional table of relationships, such as similarity and difference, among pairs of elements in the genomes, its basic idea is to present the pairwise data matrix in a circular ideogram layout with ribbons that connects related elements, see Figure 4.3 for examples. Other than genome visualization, the circos figure was also adopted by Bostock et al. [20] for web visualization, and another related visual metaphor that shares similar characteristics is the contingency wheel [4].

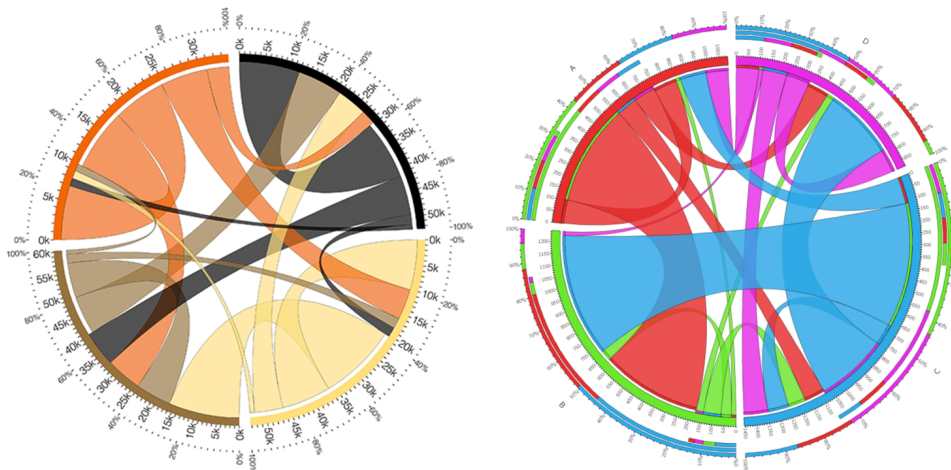


Figure 4.3: Example circos figures developed by Krzywinski et al. [74] for examining the mutual relationship among genomes.

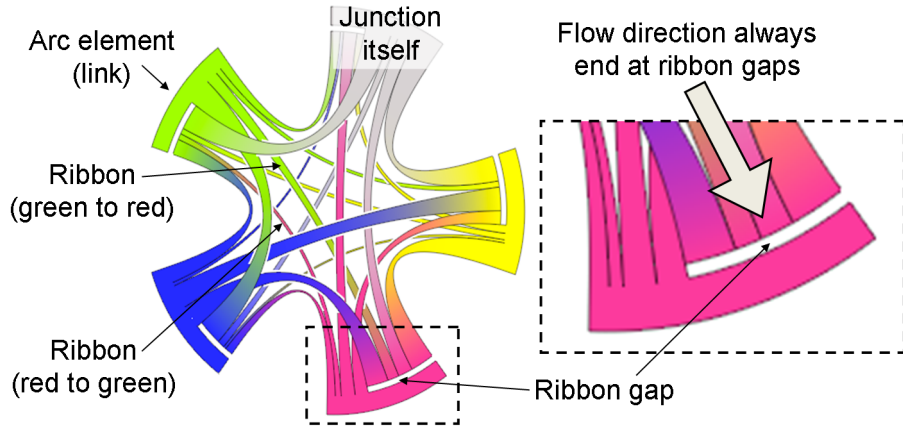


Figure 4.4: An initial design of interchange circos diagram with arc elements representing the exterior links and junction itself, and ribbons connecting arc elements representing the flows.

4.3.2 Initial Design: Interchange Circos Diagram

To develop our interchange circos diagram from circos figures, the very first step is to map the interchange information to the various visual components in a circos figure. First, we map the connecting links at a junction node (including the junction itself) as *arc elements* around the figure's boundary, and vary the angular size of these arc elements according to the total flow volume across the links, see Figure 4.4. Then, we join the arc elements with *curved ribbons* and vary the ribbon width to present the flow volume.

Moreover, we sort and render the ribbons from back to front to emphasize the flows with larger volume, and employ haloes [14, 66] to visually emphasize the occlusions between intersecting ribbons. Next, we assign a unique color to each arc element (see Section 4.3.4), and specifically assign grey to indicate the junction itself.

Lastly, since movement is bidirectional, we need two ribbons between every pair of arc elements. Thus, we highlight the ribbon direction by 1) gradually changing the color along the ribbon from its source to destination but using the source color as the dominated color, and 2) putting a *ribbon gap* (see Figure 4.4) between the ribbon and its destination arc element. Hence,

we can formulate an initial design of our interchange circos diagram as a visual representation of the interchange information at a junction, see Figure 4.4.

4.3.3 Improving Our Visual Design

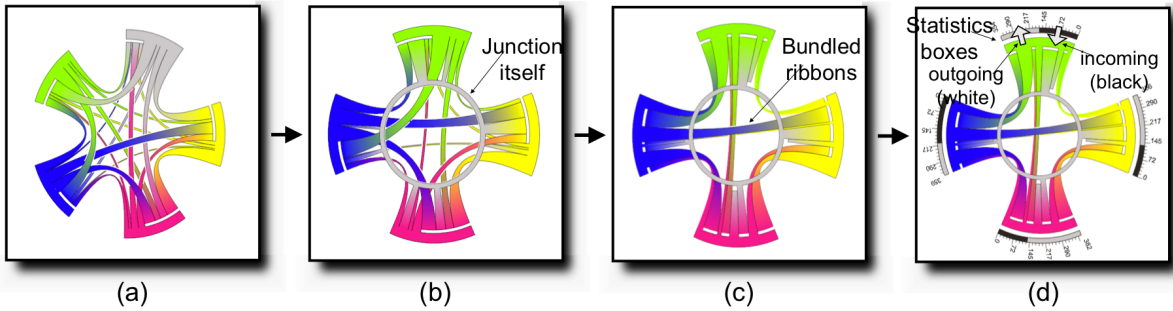


Figure 4.5: Developing the interchange circos diagram from the original circos figure: (a) the initial design in Figure 4.4; (b) use a grey-colored flyover ring (like a source/sink) for the junction itself; (c) bundle pairs of bi-directional ribbons to reduce the visual cluttering; and (d) draw white and black curved statistics boxes to present the total outgoing and incoming flow volumes.

However, this initial design still has a number of issues:

- (i) *Visual confusion.* Since the original circos figure treats all genome elements equally, it is thus natural to put the elements around the figure's circular border. Our case is, however, different because of a special link, i.e., the junction node itself. Hence, if we just present this link equally like the external connecting links, they can be mixed up, and potentially result in a visual confusion.
- (ii) *Visual cluttering.* Second, for a junction node of valence n , we have $n(n - 1)$ ribbons in total within an interchange circos diagram, e.g., the interchange circos diagram in Figure 4.5(a) has 20 ribbons. Even though we sort and render the ribbons, and apply haloes to enhance the visual occlusion, the intersecting ribbons could still be cluttered in spite of the fact that n is usually 4 or 5.

- (iii) *Visual analytics.* Lastly, domain experts may want to directly observe basic statistics in the visual representation, but such information could still be missing, or not straightforward to be seen, e.g., comparing relative flow volume between bi-directional routes.

Hence we propose the following techniques to further improve our design:

- (i) *Flyover Ring.* To address the first issue above, we isolate the junction node, i.e., the source and sink of the interchange, from the other connecting links, and use a grey-colored flyover ring to represent the junction node, see Figure 4.5(b). In this way, we can avoid the visual confusion issue as well as reduce the number of ribbons.
- (ii) *Bundling Ribbons.* To allow domain experts to visually compare the relative flow volumes between bidirectional ribbons between the same pair of links, we propose to bundle each pair of bidirectional ribbons together, see Figure 4.5(c). As for the labeled bundle shown in Figure 4.5(c), we can easily see that the blue-colored ribbon dominates; hence, there are far more people traveling from the blue to yellow link, than that of the opposite direction. In addition, this strategy can also help to address the visual cluttering problem by further reducing the total number of ribbons, e.g., from 20 in our initial design, to just the six ribbons shown in Figure 4.5(c).
- (iii) *Statistics on Flow Volume.* Lastly, we draw a pair of black and white curved statistics boxes above each arc element with angular sizes proportional to the flow volumes along the corresponding link, see Figure 4.5(d). By these statistics boxes, one can quickly identify the relative flow volume along each link. Note that we use grey to indicate the outgoing flow and black for the incoming flow, and we may also optionally put in the actual numbers of the flow volume on the boxes.

4.3.4 Coloring Arc Elements

Since there are multiple interchange circos diagrams interconnected over the underlying traffic network, see Figure 4.2(e) or Figure 4.6, we propose to improve the visual connection between them by coloring their links (and the related arc elements) with the following two constraints:

- First, links connected to a common junction node should have different colors;
- Second, a common link between two neighboring junction nodes should have the same color.

This indeed is an edge coloring problem of an undirected graph, i.e., the traffic network G . Rather than using complex combinatorial optimization, since a junction node has at most seven links (which is a very rare case), we found that it is sufficient to fulfill the above two constraints by precomputing a small number of distinct colors and then applying a simple algorithm to assign these colors to the links:

```

1: Initialize:
2: for each edge in  $G$  do
3:    $c_{ij} = \emptyset$   $\triangleright c_{ij}$  is the link color between vertex  $i$  &  $j$ 
4: end for
5: Main Loop:
6: for each edge in  $G$  (random order) do
7:    $C_i$  = colors previously assigned to links of vertex  $i$ 
8:    $C_j$  = colors previously assigned to links of vertex  $j$ 
9:    $C$  = precomputed colors -  $(C_i \cup C_j)$ 
10:   $c_{ij}$  = randomly choose a color in  $C$ 
11: end for

```

If k is the maximum vertex valency in G , the maximum number of neighboring links that any link would have is $2(k - 1)$. Hence, precomputing $2k - 1$ colors would be sufficient to fulfill the coloring constraints. In our implementation, we precompute a table of 13 colors ($k = 7$).

More than a single traffic network, we may have a series of traffic networks of different spatial scales. In this case, we should also attempt to maintain color coherence for links that exist

in networks of consecutive spatial scales. This helps to maintain the visual context when one explores across spatial scales. To address this, we first apply our color assignment method to the coarsest-scale network graph, and then progressively color the links in the next finer-scale graph with an additional constraint:

- Third, if a link exists in two consecutive network graphs of different scales, we should try to assign a similar color to its two instances.

This is done by first checking if a link exists in the previous coarser graph and retrieving its color, say c_0 , from the graph. If we need to enforce the third constraint, we assign c_0 to c_{ij} if c_0 is in C (see the main loop in pseudo code above), else we pick a color in C that is the most similar to c_0 .

4.3.5 Positioning Arc Elements

When putting interchange circos diagrams that are geographically interconnected with one another, see again Figure 4.6, we have to scale and shift (angularly) the arc elements in each interchange circos diagram because of the following two issues. First, we need to scale the angular size of the arc elements, so that angular sizes can be used to indicate relative flow volume among links in the visualization. Second, taking the interchange circos diagram at the bottom of Figure 4.6(a) as an example, we need to shift the blue arc element, so that it roughly align with the direction toward the related interchange circos diagram on the right.

To address the first issue, we first determine the junction node that has the largest sum of in and out flow volumes in the current visualization view, e.g., the interchange circos diagram at the bottom of Figure 4.6(a). Then, we constrain the angle sum of all arc elements around it to be 180 degrees, and compute the angle size of every arc element in the visualization view by a simple linear proportionality based on its flow volume. By this simple idea, we can guarantee

that the angle sum of arc elements around any node is no greater than 180 degrees, and that we can have sufficient angular space to shift the arc elements to resolve the second issue.

To further resolve the second issue to avoid overlapping the arc elements, every arc element in the current visualization is initially positioned in a way that it points toward its link direction. Then, in each interchange diagram, we simply check if any neighboring arc elements are too close to each other, and make them repel from each other. This is repeated iteratively until every pair of neighboring arc elements has a minimum gap of 10 degrees from each other.

4.4 Interface: Visualizing Interchange

This section presents our visualization interface: 1) multi-scale visualization of interchange patterns, and 2) a family of interaction techniques for exploring interchange patterns.

4.4.1 Multi-scale Visualization

As mentioned earlier, interchange patterns can emerge in different spatial scales, see Figure 4.6. Hence, given the traffic network graphs (of different scales) and the interchange matrices we precomputed from the raw trajectory data, see Section 4.2.2, we can plot the network graph associated with the current viewing scale in the visualization interface, and render the interchange circos diagrams at the visible junction nodes in that network graph.

Therefore, in case of the coarsest level (road level), we show one interchange circos diagram per road junction, and in case of region/city scales, where each partitioned area is a junction node, we show one interchange circos diagram per partitioned area and put it at the centroid of the area to avoid cluttering. See Figure 4.6 for the visualization results.

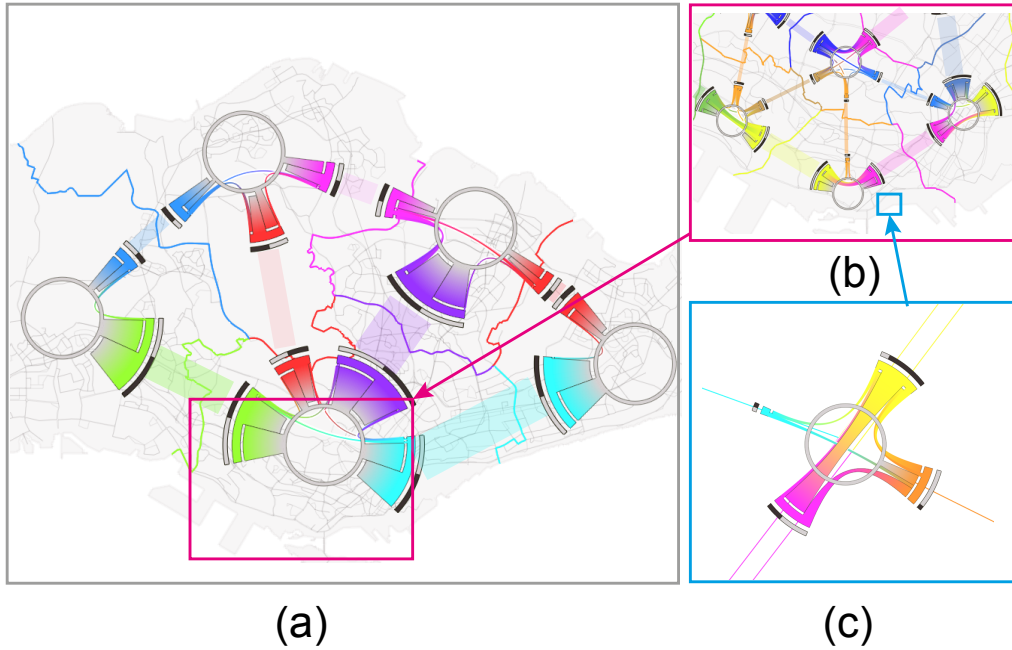


Figure 4.6: Interchange Circos Diagrams in different scales: (a) city scale; (b) regional scale; and (c) road network scale.

4.4.2 Interaction

Our system offers a family of interaction techniques to let users interactively explore the interchange patterns.

- *Select.* The user can select an interchange circos diagram by click on it. After that, the related junction/region is highlighted as a visual feedback.
- *Zoom.* If a series of multi-scale traffic network graphs is available, the user can interactively zoom in/out to examine the interchange patterns in different spatial scale. In addition, our interface also provides an interactive magnifying glass function for users to do a focus+context visualization to examine the interchange patterns.
- *Roll.* In addition, one compelling feature of our interface is that the user can roll out a series of interchange circos diagram, see Figure 4.8, and observe the temporal changes of the interchange pattern over time.

- *Time Control.* Other than rolling to see temporal changes at a junction node, the user can also interactively adjust a timer control to filter the trajectory paths against a user-preferred time interval. By this, the user can animate all interchange circos diagrams in the visualization view and observe the temporal changes.

4.5 Implementation and Results

This section presents the implementation details of our system, followed by two case studies.

4.5.1 Implementation

This system is implemented entirely in Java, so that it can run on different platforms in the future. Currently, it runs on an Intel Core i7 2 2GHz MacBook Pro with 8GB memory and an AMD Radeon HD 6490 graphics board.

Data storage. In the offline precomputation, see again Section 4.2.2, we mainly pre-compute interchange matrices for each junction node at all traffic network graphs over the partitioned time intervals. Note that we use 15 minutes as the time interval, so there are $24 \times 4 = 96$ partitioned time intervals over a day. Moreover, since there are about 1,600 junction nodes in total over all traffic network of different scales, and the interchange matrices are mostly 5×5 on average, the total memory needed to store the precomputed interchange data is around $96 \times 1600 \times 25 \times 4$ bytes, i.e. $\sim 15\text{MB}$ (note: we use 4-byte integers for the matrix elements).

Offline precomputation. Since it is impossible to load the entire raw trajectory data into the main memory, we divide the raw data into chunks and precompute the interchange matrices, i.e., $\sim 15\text{MB}$ data, for each chunk. Since interchange matrices of the same junction node can be summed, we can aggregate the overall interchange matrices for all raw trajectories by adding up matrices from the data chunks. It took about 30 minutes to preprocess one data chunk, and around 10 hours for all 20 data chunks to finish the offline preprocessing.

4.5.2 Case Study 4-1: Interchange at Metro Stations

Our system has been customized to analyze the interchange patterns at train stations in the Singapore MRT system, which consists of four metro service lines.

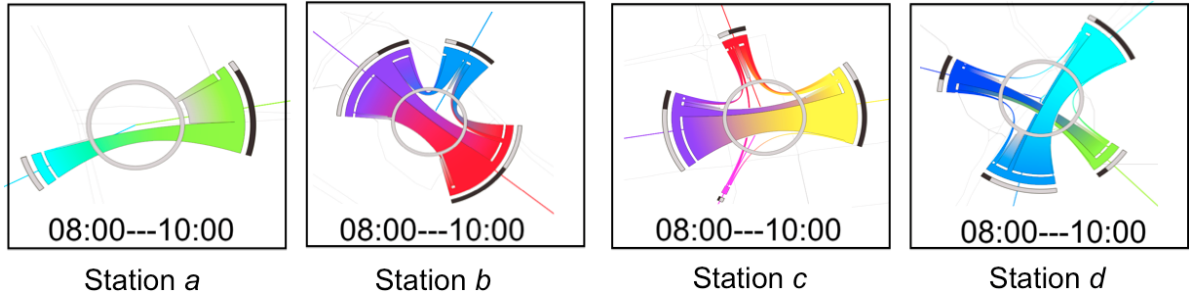


Figure 4.7: Interchange patterns at different train stations (*a-d*) in the Singapore Metro system.

As shown in Figure 4.7, we can pick a train station and visualize its interchange pattern for a user-selected time interval, which is 08:00 - 10:00 in this case. By examining these four interchange circos diagrams, we can see the relative flow volumes for different possible routes at these train stations, e.g., the major movement directions at each station as well as the relative flow volumes among the four stations. Since the selected time interval is in the morning, we can observe unbalanced flow volumes in the bundled ribbons as well as in the node-connecting links.

By using the “roll” operation, we can roll out a series of interchange circos diagrams, see Figure 4.8 to examine the temporal variations of interchange patterns at these four train stations. A common and general pattern shared by all four stations is that the most heavy traffic periods are the morning and evening peak hours, while there are far fewer commuters during the lunch hours. Specifically, we could observe the followings in our visualization:

- STATION *a* (Figure 4.8(a)) is the 2nd last station on XX MRT line. It only has two connecting links, and there are fewer traveling commuters compared to the other interchange stations as presented in the figure. In addition, our visualization can also reveal

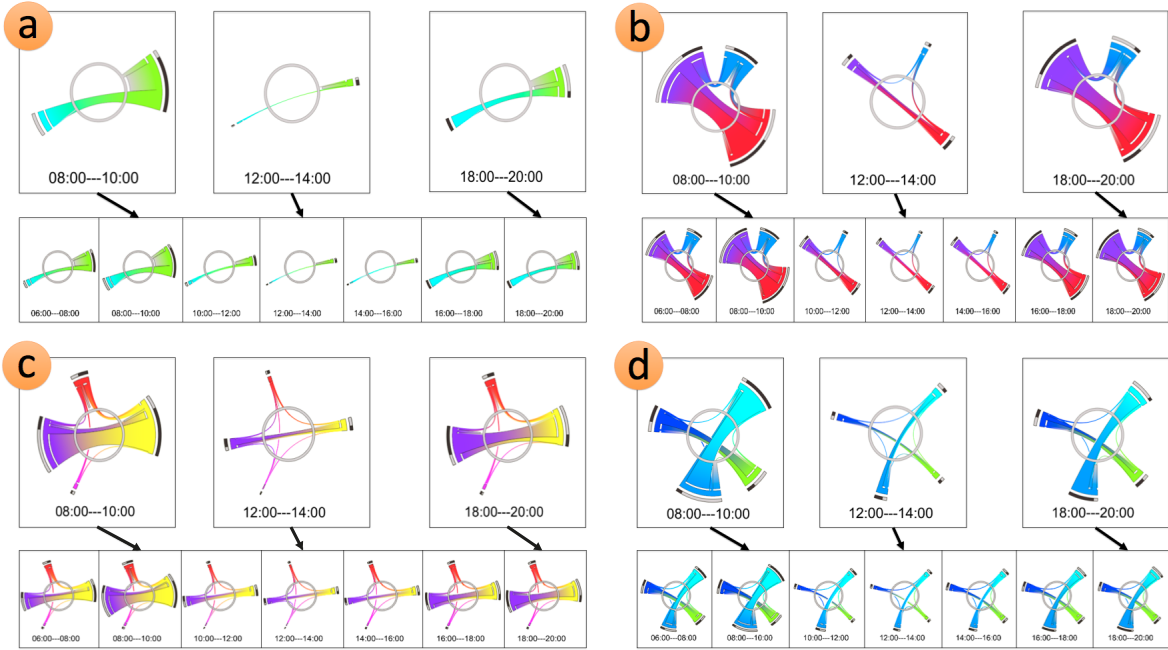


Figure 4.8: Exploring the temporal changes (over a day) in the interchange patterns at four different train stations (*a-d*) in the Singapore MRT system.

that during the morning peak hours, almost all commuters who enter this station come from the east side, and these people are almost equally distributed when they leave this station: roughly half of them continues their journey to the next station while the other half goes into the dominion of the station. And when evening comes, the interchange pattern basically reverses.

- **STATION *b*** (Figure 4.8(b)) is a busy interchange station connecting XX and YY MRT lines. Comparing its temporal variation against that of the other three stations, we can clearly see that its flow volume is always larger than that of the others. Moreover, commuters who enter this station from YY line on the top are (always) nearly equally redistributed into the east and west connecting directions on XX line, and XX line is usually busier than YY line. Lastly, we can also observe asymmetric flow volumes between the east and west connecting directions in this station during the morning and evening periods similar to that in STATION *a*.

- STATION *c* (Figure 4.8(c)) is an interchange station between XX and WW MRT lines with four connecting links. As seen from the figure, XX line basically dominates the flow in this station. Though relatively fewer commuters on XX line transit to WW line here, commuters to or from WW line appear to redistribute fairly equally for all different outgoing routes in this station, showing that the newly-established line, i.e., WW line, is like a supporting branch with line 1 being the main route.
- STATION *d* (Figure 4.8(d)) is an interchange station linking ZZ and WW MRT lines. Interestingly, we find that the traffic flow volumes across the two lines are nearly the same, but these two service lines are relatively independent of each other, i.e., relatively not too many commuters transit between them, as compared to XX and ZZ line commuters in STATION *c*.

4.5.3 Case Study 4-2: Intersection Capacity Utilization

The intersection capacity utilization (ICU) method [65] is a standard way in transport research to measure the utilization rate of a road junction.

Our interface can also be used to estimate ICU at road junctions because one key factor that affects ICU is the relative amount of incoming and outgoing flow volumes from each direction at the road junction. Basically, the more balanced the flow volumes at different connecting links are, the junction will usually have a higher ICU rating.

The left and right hand sides of Figure 4.9 compare lower and higher ICU ratings, respectively, at a road junction during different time periods. Figure 4.9(a) has a lower ICU rating since the traffic flows from yellow to violet dominate the junction utilization; moreover, both the orange and yellow connecting links are highly unbalanced. Figure 4.9(b) has a higher ICU rating because traffic flows from each direction, as well as the incoming/outgoing flow volumes are

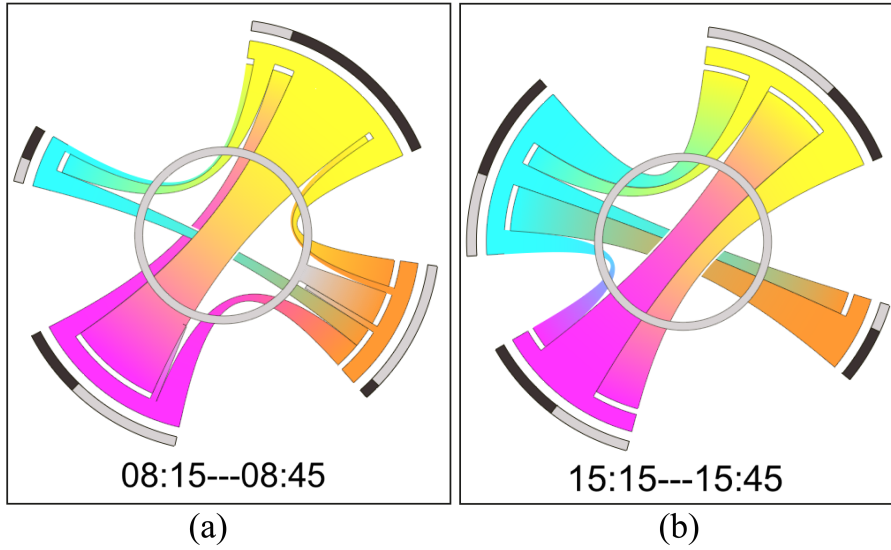


Figure 4.9: Comparison of lower (left) and higher (right) ICU ratings at a road junction during different time periods. (a) The traffic flow from yellow to violet dominates the junction utilization; moreover, both the orange and yellow connecting links are highly unbalanced. (b) Traffic flows from different links in the junction are fairly balanced and the incoming/outgoing flows for each connecting links are also fairly balanced.

relatively more balanced. With our interface, domain experts can efficiently identify potential road junctions with low ICU rating across different time of the day.

4.6 Discussion

This chapter explores and analyzes interchange patterns emerged from the EZLink data. Interchange pattern can be considered as a local aspect of OD pattern, which associates with only the trajectories passing through a specific position in the transport network. In particular, we have proposed a novel visual representation, i.e., the *interchange circos diagram*, to intuitively and effectively present the pattern.

On contrast, many existing visualization methods represent traffic flows by considering locations (junction nodes) in a pairwise manner. They aggregate the trajectory flows by computing only the total flow volume between every pair of neighboring nodes, and present these aggre-

gated information as (bidirectional) arrows with varying width and color to show the corresponding flow volume. Such approach is intuitive and has been adopted in many applications, but it is not sufficient to reveal the interchange patterns because the interchange information has been lost when aggregating data.

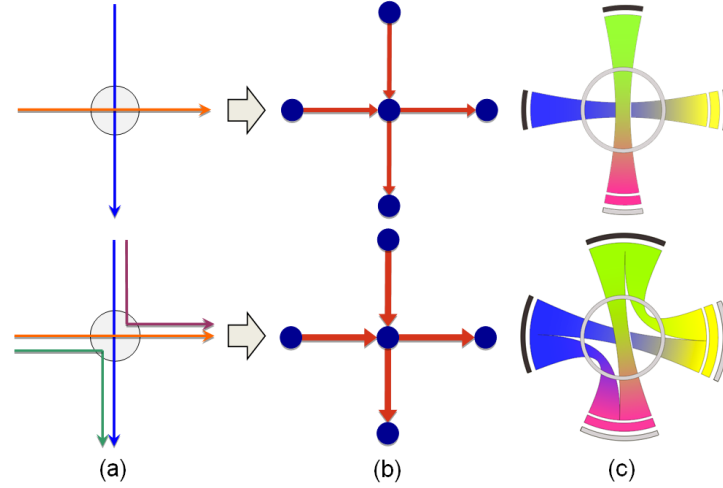


Figure 4.10: Comparing interchange circos diagrams with existing visualization approach. (a) Two sets of raw trajectories; (b) Existing approach aggregates flows between pairs of locations and draws arrows to indicate the aggregated flow volume; (c) Our interchange circos diagrams are able to reveal the detail on the interchange patterns.

Figure 4.10 compares interchange circos diagrams with the existing visualization approach. Here we show two simple examples that contain two and four trajectories, see top and bottom of Figure 4.10(a), respectively. After the data aggregation, both sets of trajectories result in a very similar aggregated visualization, see Figure 4.10(b), but in sharp contrast, our interchange circos diagrams are able to present to us clearly the difference in the interchange patterns emerged from the two trajectory sets, see Figure 4.10(c).

Scalability of our method. Since our visualization interface works with the precomputed interchange data, we do not need to load the raw trajectory data in the program run-time. Hence, it is independent of the amount of raw trajectories. However, it does depend on the time resolution we choose and the number of junctions we have in the traffic network graph because they affect the size of the precomputed interchange data.

Chapter 5

Visual Analytics for Waypoints-Constrained OD Patterns

Origin-Destination (OD) pattern is a highly useful means for transportation research since it summarizes urban dynamics and human mobility. However, existing visual analytics are insufficient for certain OD analytical tasks needed in transport research. For example, transport researchers are interested in path-related movements across congested roads, besides global patterns over the entire domain. Driven by this need, we propose *waypoints-constrained OD visual analytics*, a new approach for exploring path-related OD patterns in an urban transportation network. First, we use hashing-based query to support interactive filtering of trajectories through user-specified waypoints. Second, we elaborate a set of design principles and rules, and derive a novel unified visual representation called the *waypoints-constrained OD view* by carefully considering the OD flow presentation, the temporal variation, spatial layout, and user interaction, etc. Finally, we demonstrate the effectiveness of our interface with two case studies and expert interviews with five transportation experts.

5.1 Introduction

Origin-Destination (OD) pattern is a fundamental concept in transportation, summarizing people and vehicle movements across geographical regions [129]. Studies show that analyzing OD patterns can facilitate the understanding of urban dynamics and human activities, e.g., estimating region functionality [146], revealing urban structure [68], and studying congested road usage [132].

As such, OD pattern has been an important topic in the study of transportation and urban planning. However, visualizing OD patterns has always been challenging. First, considering real transportation data with numerous locations and passenger trajectories, huge amount of OD pairs could be easily produced. As a result, the visualization will likely end up with visual clutter if we simply employ conventional visualization methods like flow map. Second, recent research by Wang et al. [132] shows that only a few (less than 2%) of the road segments in urban areas give rise to congestion. This motivates transportation researchers to study OD patterns subject to specific locations/paths rather than to the entire city. However, existing visual analytic methods generally focus on global OD flows across regions and ignore OD flows constrained along specific locations/paths. Moreover, city-scale OD patterns can be highly complex. That is, the traffic condition in a city could have huge spatial and temporal variations, e.g., peak versus non-peak hours, busy versus deserted roads, etc. Lastly, transportation researchers are concerned with not only the OD flow volumes, but also the movement paths of the OD flows.

To address the above spatial-, temporal- and path-related requirements, we design a new visual analytics approach, namely *waypoints-constrained OD visual analytics*, aiming to help users analyze OD patterns associated with trajectories that successively pass through specific links or waypoints in the transportation network. This approach could help transportation researchers in transportation planning and traffic management, e.g., in a situation where some subway

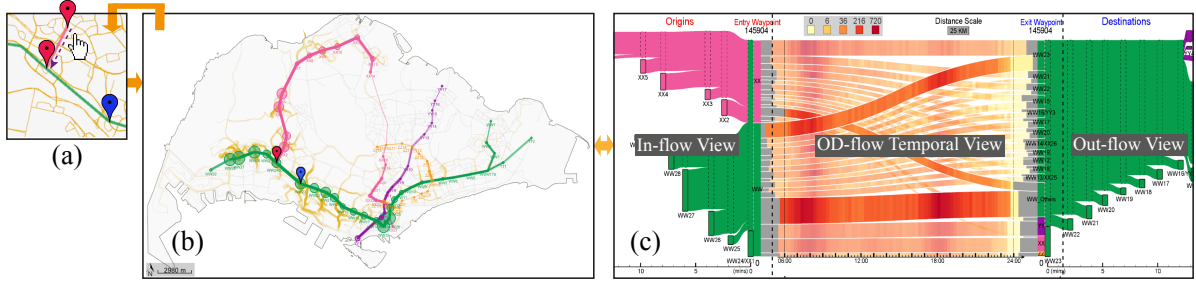


Figure 5.1: Overview of our waypoints-constrained OD visual analytics interface. We can interactively (a) select and modify the entry (red) and exit (blue) waypoints in the transportation network, (b) filter passenger trajectories that successively pass through the selected waypoints, and (c) present the waypoints-constrained OD view to support the analytical tasks.

routes are disrupted, service providers can identify closely-connected origins and destinations and provide emergency bus services for commuting the passengers.

Our approach is achieved through an iterative design process. First, we set forth the requirements and analytical tasks in collaboration with the transportation researchers. Second, we elaborate a set of design principles and rules, and carefully consider the OD flow presentation, the temporal variation, spatial layout, and user interaction, etc. when designing the *waypoints-constrained OD view*. Third, we use a hashing-based query method to support interactive filtering with over ~ 2.1 millions of daily passenger trajectories. Lastly, to demonstrate how our visual analytics interface helps to study and explore the Singapore public transportation data, we present two case studies and conduct an expert review with five transportation researchers.

5.2 Overview

Transportation Data. The data we employed is from the Singapore Mass Rapid Transit (MRT) system, which is a metro system consisting of about 2.1 million daily passenger trips. In Singapore, passengers carry personalized RFID cards to enter and leave the public transportation system by tapping their own RFID cards on the card readers available in the stations. The card readers can automatically record various trip information such as card ID, tap-in time, tap-out

time, related stops, etc. From this raw data, one can reconstruct the passenger trajectory path with time stamps over intermediate stops for every trip record [38].

Basic Concepts. The public transportation network can be represented as a directed graph $G := (V, E)$, where V is a set of nodes in G and E is a set of directed edges connecting neighboring accessible nodes (locations). Hence, a trajectory T is a sequence of consecutive directed edges in G :

$$T := v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_m,$$

where $v_i \in V$ and $2 \leq m \leq |V|$. Moreover, we have a timestamp t_i at each v_i along trajectory T . The waypoints-constrained OD pattern associates with trajectories that successively pass through two user-specified waypoints in the transportation network: the *entry waypoint* node, which receives passengers coming from different origins, and the *exit waypoint* node, which sends passengers to their destinations. As a convention, we represent the entry and exit waypoint nodes as red and blue glyphs, respectively, see Figure 6.1(a) & (b).

Analytical Tasks. In our collaboration with transportation researchers, we identified a family of analytical tasks. First, our interface should support interactive filtering of trajectories, say $\{T\}$, that successively pass through the user-specified entry and exit waypoints for a given time period. Then, the interface should present spatial- and temporal-related information to support the following basic tasks:

- *T5-1*: Find the origins and destinations from $\{T\}$;
- *T5-2*: Examine and compare the flow volumes among the OD pairs derived from $\{T\}$;
and
- *T5-3*: Examine and compare the temporal changes in flow volumes among the OD pairs.

Besides, they would also like to perform some path-related tasks specifically for OD patterns in urban traffic data:

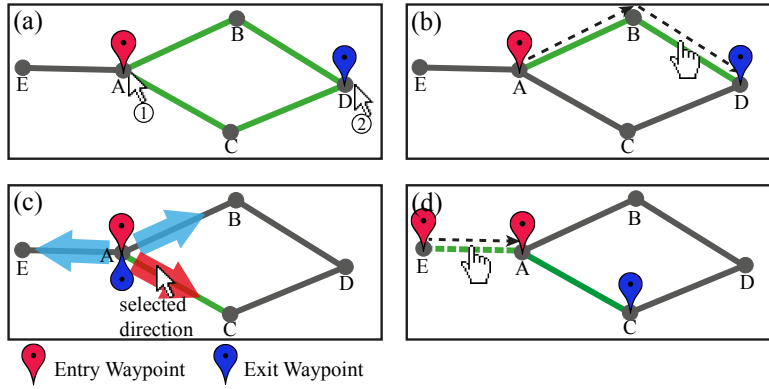


Figure 5.2: Interactive waypoints specification: a) mode 1: two successive clicks to select an entry and then an exit waypoint; b) mode 2: drag to define a path from the entry waypoint to exit waypoint; c) mode 3: click to select an entry waypoint and an outgoing direction (red arrow); and d) click to select and then drag to modify an existing waypoint.

- *T5-4*: Present the paths through which the trajectories go from origins to destinations; in some situations, we may have multiple paths between the entry and exit waypoints.

System Overview. Figure 6.1 shows the workflow. First, the user interactively manipulates the entry and exit waypoints by simply clicking and dragging the red and blue glyphs in the network (Figure 6.1(a)). Upon changes in waypoint locations or user-specified time period, the interface automatically filters and queries relevant trajectories from millions of daily trajectories, and presents a map view of the retrieved trajectories in a split second (Figure 6.1(b)). Then, the user can bring in the *waypoints-constrained OD view* (Figure 6.1(c)) to explore not only the spatial and temporal semantics of the OD patterns, but also the path-related information. Note that this *waypoints-constrained OD view* has three main components: *in-flow view*, *OD-flow temporal view*, and *out-flow view*, see Section 5.4 for detail. These components work together to support the analytical tasks above.

5.3 Waypoints-Constrained OD Query

5.3.1 Interactive Waypoints Specification

To construct a visual query, users can interactively specify and manipulate the (entry and exit) waypoints by simple mouse click and drag with the following three modes:

Mode 1: two successive mouse clicks to select an entry waypoint and then an exit waypoint, say A and D in Figure 5.2(a). Our interface then applies the query method in Section 5.3.2 to retrieve all relevant trajectories through A and then D by considering all possible time-efficient paths from A to D , i.e., $A \rightarrow B \rightarrow D$ and $A \rightarrow C \rightarrow D$.

Mode 2: click to select an entry waypoint (A) and then drag along the network to define a path (ABD), see Figure 5.2(b). The node at which the mouse button is released defines the exit waypoint (D), and we consider only the trajectories along the dragged path $A \rightarrow B \rightarrow D$ but not $A \rightarrow C \rightarrow D$.

Mode 3: long-press to select a common entry and exit waypoint, say A in Figure 5.2(c). The interface then shows all outgoing (blue) arrows emerged from A . In this mode, the user can select different outgoing directions from A and explore the OD patterns of trajectories along different directions from the same junction node.

Once the waypoints are specified, the user can interactively modify them on the map, see Figure 5.2(d). In case the entry and exit waypoints become coincident, the query mode smoothly changes from mode 1/2 to mode 3, and vice versa.

5.3.2 Hashing-Based Trajectory Query

To support such interactive query, we need to efficiently filter out relevant trajectories against a given time period (Δt). Rather than scanning through the nodes along every trajectory in the

data, we use a hashing-based method. First, we attach a unique ID (tid) to each trajectory, and define 72 equal time intervals from 6 A.M. till midnight, each covering 15 min. The choice of this interval size is driven by a common practice by transportation researchers: when modeling and analyzing transport data, they normally set a minimum analysis time interval, which should not be too short, so that there are sufficient samples in each interval, and should not be too long to avoid losing the details. Altogether, there are two stages in the query method:

Indexing Scheme (offline). First, for each edge $e := \langle v_i, v_j \rangle$ in G , we record per time interval a list of tid of trajectories that pass through e . A trajectory T is said to pass through e within Δt if T 's $[t_i, t_j]$ overlaps Δt , where t_i and t_j are time at which T passes through v_i and v_j of e , respectively.

Second, we build a hash table for each node v in G , where the hash key is a tid and the hash value is the corresponding time at which the trajectory passes through v . Since hashing can be done in $O(1)$ time, we can quickly check if a given trajectory (tid) passes through v , and if this is the case, we can also obtain the related timestamp (t_v).

Trajectory Query (online). Given entry waypoint A , exit waypoint D , and time interval Δt , our interface performs the following two steps to extract the relevant trajectories.

In the first step, our goal is to determine a set of candidate trajectories (i.e., a list of tid 's), say S_A , that exit from A within Δt . After identifying A 's outgoing edges that are relevant to the query, we retrieve and combine the lists of precomputed tid 's through each edge for the 15-min. time interval(s) covered by Δt . Note that if Δt covers more than one 15-min. time interval, we have to remove duplicated tid 's when combining the precomputed lists since some trajectories may appear in two consecutive time intervals. It is because when a passenger travels through a link (edge) in the transport network, he/she may start the travel within a certain time interval and end it within the next time interval.

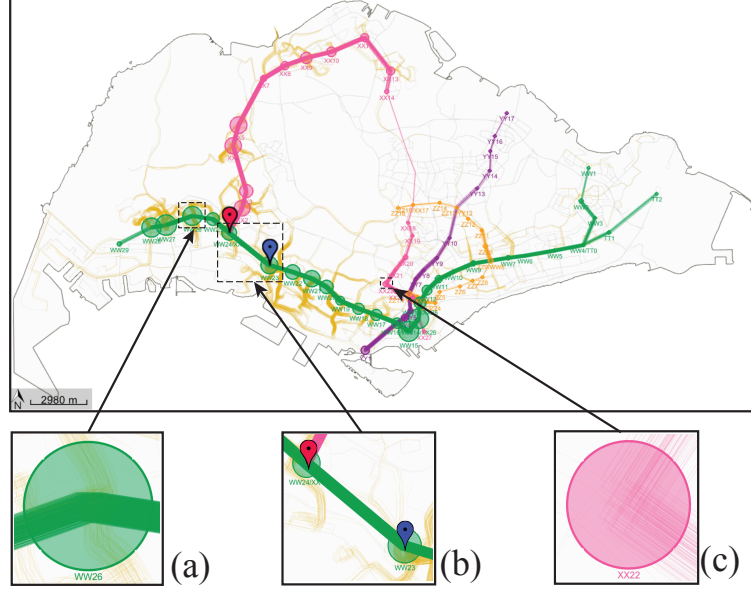


Figure 5.3: The map view overviews the waypoints-constrained trajectories: a) an origin; b) red and blue glyphs as entry and exit waypoints, resp.; and c) a destination.

In the second step, our goal is to filter the trajectories in S_A , and output those that pass through D within Δt . Here we employ the precomputed hash tables for speedup. In detail, we first retrieve the hash table at exit waypoint D , and use it to remove the trajectories (in S_A) that do not pass through D , or pass through D but outside Δt (by the hash value). In case of mode 2, we need to perform this test for every intermediate node along the user-selected path from A to D .

After these tests, the remaining trajectories are the query result for constructing the OD flows. Note also that in case of mode 3 (with same entry and exit waypoints), we can skip the second step and output S_A as the result.

5.3.3 Map View

After the query, we present a map view (Figure 5.3) to overview the amount of retrieved trajectories over different segments in the transportation network. This map view provides intuitive

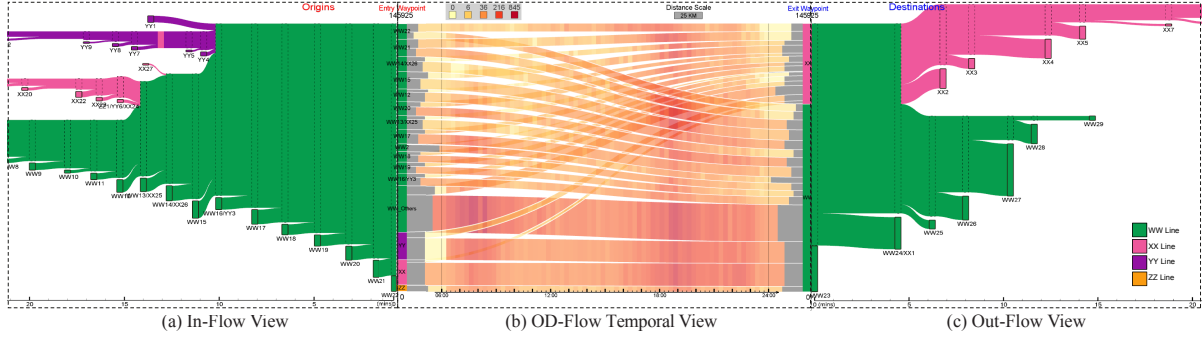


Figure 5.4: An example *waypoints-constrained OD* view with three components: (a) in-flow view, (b) OD-flow temporal view, and (c) out-flow view, as an integrated and coordinated solution to support the visual analytical tasks.

spatial information essential for locating the origins and destinations in the physical space (*Task T5-1*).

This map view is created by hardware rendering. First, we randomly jitter the position of each trajectory by a few pixels and render each of them with low transparency, so the resulting plotting effect can roughly reveal the flow volume. Note also that we follow the Singapore MRT (metro) coloring scheme to color different parts of the trajectories, e.g., green for WW line and red for XX line.

Besides, the map view also presents: *Entry and Exit Waypoints* as red and blue glyphs, respectively, on the map, see Figure 5.3(b). *Origins and Destinations* as hollow circles positioned at their locations with radii revealing the corresponding flow volume, see the hollow circles in Figure 5.3(a) & (c).

5.4 Waypoints-Constrained OD View

In this section, we first discuss the design philosophy behind our interface. Then, we elaborate the three component views in the waypoints-constrained OD view, and discuss alternative designs. Lastly, we present the layout algorithm and user interaction we developed in the interface.

5.4.1 Design Philosophy

By the map view, we can present aggregated flow volumes across origins and destinations, and support analytical task *T5-1*. However, it is clear that the map view alone is insufficient for other tasks: *T5-2* to *T5-4*, e.g., Figure 5.3 does not allow us to examine and compare flow volumes among different OD pairs from WW28 (Figure 5.3(a)), so it cannot handle *T5-2*.

This calls for a new visualization design to address tasks *T5-2* to *T5-4*. In particular, we aim to depict the OD pairs (*T5-2*), the temporal-related information (*T5-3*), and the path-related information (*T5-4*) in an integrated and coordinated fashion. This is a very challenging problem: i) visual clutter could easily occur given excessive OD pairs; ii) integrating temporal-related information of OD pairs in the visualization could further increase the visual clutter; and iii) it is nontrivial to also support the tracing of flow paths from origins to destinations. To address these challenges, we identify the following principles as guidance of our design:

- **Overview+Detail.** The visualization should support an overview of the OD patterns, e.g., appropriately summarizing the origins and destinations, in order to address the visual clutter issue when presenting the OD pairs. Certain interactive exploration should also be incorporated in the design to allow users to further analyze the OD patterns with controllable amount of details on demand.
- **Visual Correlation with Transport Semantics.** The visualization should reveal the semantics of the transport network to promote the correlation between elements in the visualization and the actual objects they represent, e.g., MRT stations and lines. Since the visualization has to include OD pairs, temporal- and path-related information, we cannot directly put the design on the map view.
- **Intuitive Spatial Layout.** The spatial layout in the visualization should be intuitive for users to explore the origins, the destinations, and the OD flows in-between them.

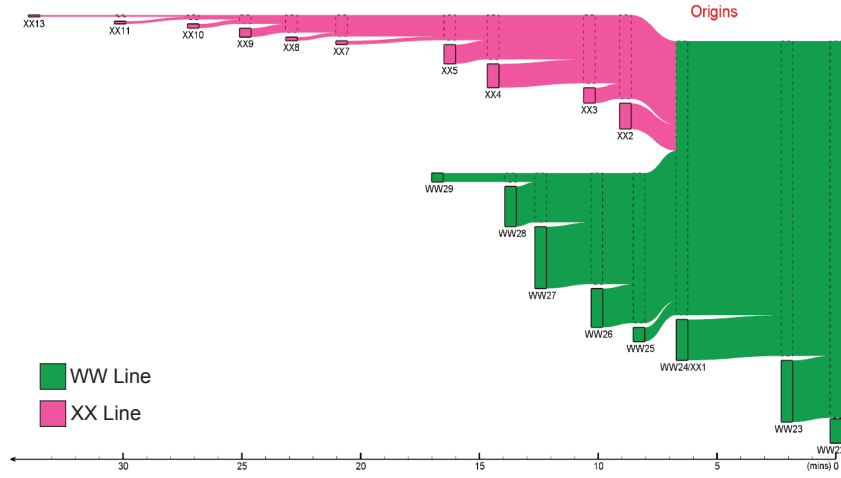


Figure 5.5: An example in-flow view: vertical boxes show the origins with heights to indicate flow volumes, and ribbons to show flow aggregation before the entry waypoint.

Furthermore, the spatial layout should facilitate intuitive user interaction needed by the users.

By considering the above design principles, we formulate a novel visual design, namely the *waypoints-constrained OD view* with three component views: in-flow, OD-flow temporal, and out-flow views (Figure 5.4(a-c)). In particular, we establish the following convention rules to meet the “*visual correlation*” and “*spatial layout*” principles:

First, we separate the three component views by two vertical bars (Figure 5.4) that represent the entry and exit waypoints: origins on the left (in-flow view), destinations on the right (in-flow view), whereas the connections between OD pairs are in the middle (OD-flow temporal view). Hence, this integrated design can naturally present the OD flows, which generally go from left to right. Second, the vertical dimension in the three component views always indicates the flow volume. Third, we always try to use the standard Singapore MRT coloring scheme for the visual elements in the visualization since this helps reveal the semantics of the transport network, e.g., green for the WW line, red for the XX line, purple for the YY line, yellow for the ZZ line, etc.

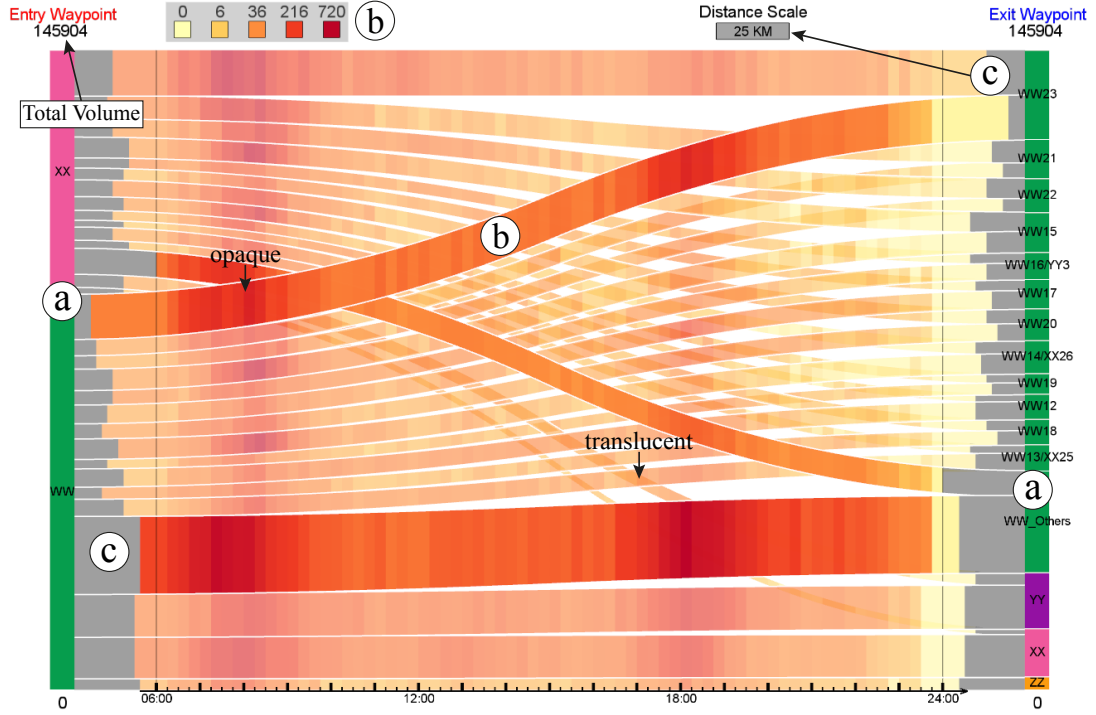


Figure 5.6: An example OD-flow temporal view that presents the OD patterns for the trajectories shown in Figure 5.3: (a) the two vertical bars represent the origin waypoint (left) and destination waypoint (right), and the colored boxes along the two bars represent the corresponding origins and destinations, respectively, (b) ribbons connect the origin and destination boxes and show the OD flow volumes in-between the corresponding OD pairs with embedded heat maps to reveal flow volume variation over time, and (c) relative (average) travel distance of each OD pair. Note also that the heat maps are colored by the color bar on top, e.g., the rightmost deep red is used for flow volume in range (216, 720], etc.

5.4.2 The In-Flow and Out-Flow Views

These two views are designed for users to effectively compare flow volumes of different stations (part of *Task T5-1*) and visualize the paths from origins to destinations (*Task T5-4*).

First, origins and destinations are presented as solid rectangular boxes; their colors follow the Singapore MRT coloring scheme, e.g., the green and red boxes in Figure 5.5, while their heights indicate the associated flow volume. Second, the horizontal axis in both views (Figure 5.5 (bottom)) indicate the travel time from origins to entry waypoint (in-flow view), or from exit waypoint to destinations (out-flow view). So, we can easily observe and compare the travel

time between different origins/destinations and the two waypoints. Third, we adopt the Sankey flow diagram to connect the boxes and their associated waypoint with smooth ribbons whose heights indicate the (aggregated) flow volume.

From the in-flow view in Figure 5.5, we can see that stations WW23 and WW27 contribute the most in terms of flow volume to entry waypoint WW22. Besides, flows from the XX line merge into the green WW line at station WW24/XX1, indicating that passengers from the XX line transfer to WW line at WW24/XX1. Here, different portions of height vertically along station WW24/XX1 reveal the aggregated flow volumes from the XX line, from the WW line (WW25 to WW29), and from WW24/XX1 itself. Note that in Singapore, interchange stations connecting multiple MRT lines compose of multiple IDs, e.g., WW24/XX1 is the 24th station along the WW line and 1st station along the XX line.

5.4.3 The OD-Flow Temporal View

The OD-flow temporal view (Figure 5.6) is designed to support *Task T5-2* & *Task T5-3* with the following visual elements:

- (i) *Origins and Destinations.* The two long vertical bars represent the entry and exit waypoints (Figure 5.6(a)) and contain boxes that represent the origins and destinations. These boxes can be manipulated by users for “*overview+detail*” exploration, i.e., users can overview the OD flows by aggregating the origin/destination boxes, or explore their details by decomposing them, see Section 5.4.6. Note also that we group the boxes by MRT lines and sort them by the method in Section 5.4.5.
- (ii) *OD-Pair Flow.* To address *Task T5-2*, we use smooth ribbons (Figure 5.6(b)) to connect the origin and destination nodes (boxes) to show the flows for each OD pair over the given time period shown on the bottom of the view. Here, we emphasize the OD flows with

larger flow volumes by rendering the ribbons from back to front in ascending order of flow volumes, and we add halos around the ribbon boundaries to help reveal the layering.

- (iii) *Temporal Variation of OD-Pair Flow.* To address *Task T5-3*, we adopt a heat map visualization on each ribbon to present the associated temporal variation of flow volume over the given time period. In detail, we horizontally divide each ribbon into column segments, each corresponding to a 15-min. time interval along the horizontal time axis on the bottom. Each column segment is then colored based on the color map shown on top of the view.

We consider two mechanisms to render overlapping ribbons in the OD-flow temporal view. By default, we employ translucency to blend these ribbons since this approach can preserve the visual connection and continuity of ribbons, enhancing the tracing of ribbons on the back layers. Other than that, users can optionally make the ribbons opaque, so that interested ribbons can be highlighted, e.g., the front-most ribbons in Figure 5.6.

- (iv) *OD-Pair Travel Time/Distance.* Optionally, users can examine and compare the relative (average) travel time/distance among the OD pairs by looking at the gray segments at the ends of the ribbons near the two vertical bars. For the case shown in Figure 5.6(c), the gray segments show the travel distance of the OD pairs, see also the time/distance scale mark on top of the view for facilitating the visual examination and comparison.

- (v) *Multiple Inter-Waypoints Paths.* In some situations, we could have multiple time-efficient paths in between the entry and exit waypoints, e.g., the two green paths in Figure 5.2(a). In this case, we put in additional vertical boxes in the middle of the OD-flow temporal view to bundle the OD flows across different time-efficient paths (see the green and purple bars in the middle of Figure 5.9(a & b)). This complements the in- and out-flow views to help users explore the path details.

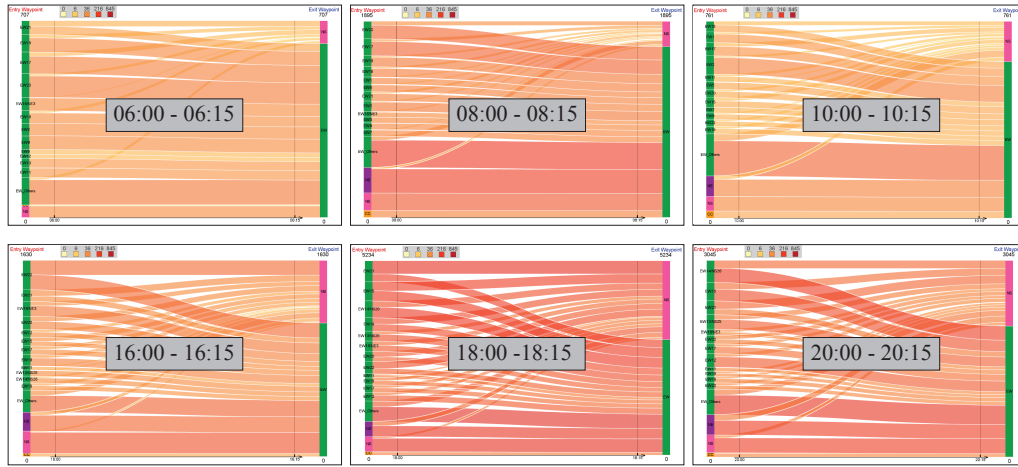


Figure 5.7: Design alternative: small multiples that show flow volume difference over six different time periods with the same entry and exit waypoints as in Figure 5.4. The individual diagrams are too small to present useful information.

Figure 5.6 shows an example OD-flow temporal view for the trajectories that pass through the entry waypoint (station WW24/XX1) and the exit waypoint (station WW23) during the time period 6:00 to 24:00. The origin WW stations (Figure 5.6(a) left) are aggregated as one single group, while the destination WW stations (Figure 5.6(a) right) are disaggregated as twelve single stations and one sub-group. The visualization here shows that there is slightly higher incoming flows originated from the WW stations than from the XX stations, and most trajectories end at WW stations. Moreover, we can observe the morning and evening peak patterns by looking at the heat maps embedded on the ribbons.

5.4.4 Discussion: Design Alternatives

This subsection discusses design alternatives for *T5-2* to *T5-4*:

To support task *T5-2*, which focuses on flow volumes among OD pairs, we connect origins and destinations with Sankey-style ribbons (Figure 5.4(b)). Such a design allows for hierarchical clustering of nodes with different details, thus promoting the “*overview+detail*” principle, yet most existing OD visualization methods do not exhibit this flexibility.

To support task *T5-3*, which focuses on temporal changes of flow volumes, we embed heat maps along the ribbons in the OD-flow temporal view (Figure 5.4(b)). Heat map is a widely-accepted and well-recognized method for depicting variations of data over time, and has been applied and proven to be efficient in visualizing temporal OD data [22].

Typical alternative designs are small multiples and animations. Small multiples can be constructed by putting together multiple OD-flow temporal views, each for a different time interval, see Figure 5.7. However, due to the small size of individual views, it is difficult to examine and compare flow volumes of OD pairs across different time periods. For instance, we can intuitively observe morning and evening peak flows in Figure 5.4(b) (with 72 time periods) but not in Figure 5.7 (with only six periods, when occupying a similar size as Figure 5.4(b) in the paper). On the other hand, animations have been shown to be useful in some cases, but they are not effective for supporting statistical analysis [103].

To support task *T5-4*, i.e., to present the paths through which the trajectories go from origins to destinations, our visualization adopts Sankey-style ribbons to support quantitative flow tracing across trajectory paths [101], see Figure 5.4(a & c).

Existing OD visualizations can be summarized into three categories based on the amount of path-related information they present: i) absent flow paths, e.g., OD matrix visualizations without any intermediate location [51, 142]; ii) discontinuous flow paths, e.g., flow maps with arrows connecting neighboring locations [50, 10]; iii) flow paths that are continuous but not existing in reality, e.g., bundled flow maps [98, 127]. None of these methods can support well the flow tracing from *multiple* origins to *multiple* destinations. Compared to these methods, our in-flow and out-flow views, which integrate with the OD-flow temporal view, can intuitively reveal how the flows aggregate before reaching the entry waypoint and how the flows distribute after leaving the exit waypoint.

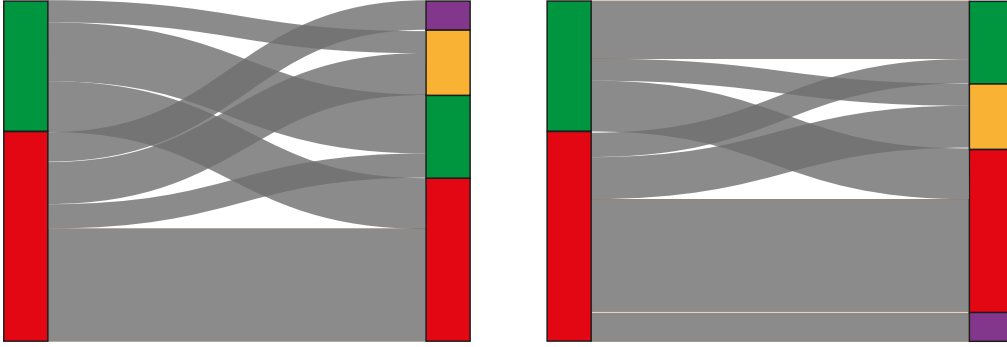


Figure 5.8: Illustration: effects of ordering origin and destination boxes by flow volumes (left) and by overlapping minimization (right) in the OD-flow temporal view.

5.4.5 Ordering of Origin/Destination Boxes

The readability of the OD-flow temporal view (see again Figure 5.6) is highly affected by the visual clutter among the ribbons. This problem is commonly found in node-link graphs, where several edge crossings reduction algorithms [59, 33] have been proposed to reduce the visual clutter by repositioning the nodes. Here, we adopt the randomized method in [126] to reorder the origin and destination boxes on the left and right vertical bars, but instead of reducing the number of edge crossings, we aim to minimize the overlapping area among the ribbons since the ribbons in our visualization are much wider than the edges in conventional node-link graphs.

In detail, our heuristic method starts by ordering the origin and destination boxes in ascending order of flow volumes, see Figure 5.8(left). Then, we randomly shuffle the boxes on each vertical bar until the total overlapping area cannot be further reduced. Lastly, we select the solution with least total overlapping area. Note also that we keep the grouping of origin and destination boxes based on MRT lines, so the re-ordering is applied only among groups or nodes with the same parent in the hierarchy. Figure 5.8(right) presents a typical result, showing that the total overlapping area (dark gray) can be effectively reduced by our method.

5.4.6 User Interaction

Furthermore, we offer a set of user interaction methods to facilitate “*overview+detail*” exploration of OD patterns apart from waypoints specification and time period selection:

Aggregate/Disaggregate the origins/destinations that belong to the same group, e.g., the same MRT service line. This allows users to interactively control the number (and details) of OD pairs to be displayed and to explore the OD pairs at different levels of detail on demand.

Filter the origins/destinations and their corresponding OD-pairs, enabling users to remove less interested OD pairs and concentrate on the remaining OD flows.

Highlight the OD pairs by selecting and emphasizing interested ribbons. This action brings the selected ribbons forward in the layering order and makes their colors opaque (see the front-most ribbons in Figure 5.6(b)).

5.5 Evaluation and Discussion

5.5.1 Performance Evaluation: Trajectory Query

We evaluated the performance of the hashing-based trajectory query method on a PC with a dual Intel(R) Xeon(R) E5-1650 CPU and 16 GB memory, and used the Singapore MRT transport data with two million passenger trajectories on a typical working day. The trajectories contain 8.3 stops on average from origin to destination, so the whole data contains around 16.5 million trajectory nodes in total.

In this experiment, we compare the performance of our method against a simple method that sequentially looks through all the nodes along each trajectory to find out the trajectories that pass through the two waypoints within a given query time period. Here, we randomly picked 10%, 50%, and 100% of the trajectories from the whole data set to form three data sets, and

performed three tests on each of them with different query time periods: 15-min., 1-hour, and 24-hour, respectively. In each test, we randomly generate 5,000 pairs of entry and exit waypoints and record the query time for each pair. We obtained the following results:

- The query time of both methods increase (apparently linear) with the number of trajectories in the data set, but only our method is affected by the length of the query time period. The performance of our method is also affected by the entry waypoint and the query time: it takes longer query time with a busy entry waypoint (with numerous trajectories) or with peak hours.
- In the worst case (100% trajectories and 24-hour query time period), our method finishes in ~ 45.3 millisec. on average, while the simple method needs ~ 8.23 sec. Note that this performance is necessary to support interactive query of trajectories when users manipulate the entry and exit waypoints or change the query time period.

5.5.2 Study 1: Transportation Network Usage Analysis

In study 1, we aim to analyze and explore flows among OD pairs, temporal- and path-related network usage, which are mainly related to tasks *T5-2*, *T5-3* and *T5-4*, respectively.

Here, we specify two time periods, 06:00-10:00 (Figure 5.9(a)) and 16:00-20:00 (Figure 5.9(b)), to compare the network usage for the same pair of waypoints: MRT stations WW16/YY3 and ZZ1/YY6/XX24. There are two time-efficient paths in-between them: the green (WW) and purple (YY) paths, see the map view in Figure 5.9(c). Correspondingly, we put two vertical boxes in the middle of the two visualizations to represent the two paths (or branches).

To address task *T5-2* in this case study, we highlight and compare the ribbons of two different OD-pair flows in Figure 5.9(a & b): one starts from WW27 while the other starts from YY1, and both end at the XX line, see Figure 5.9(d & e). From Figure 5.9(a), we can see that WW27

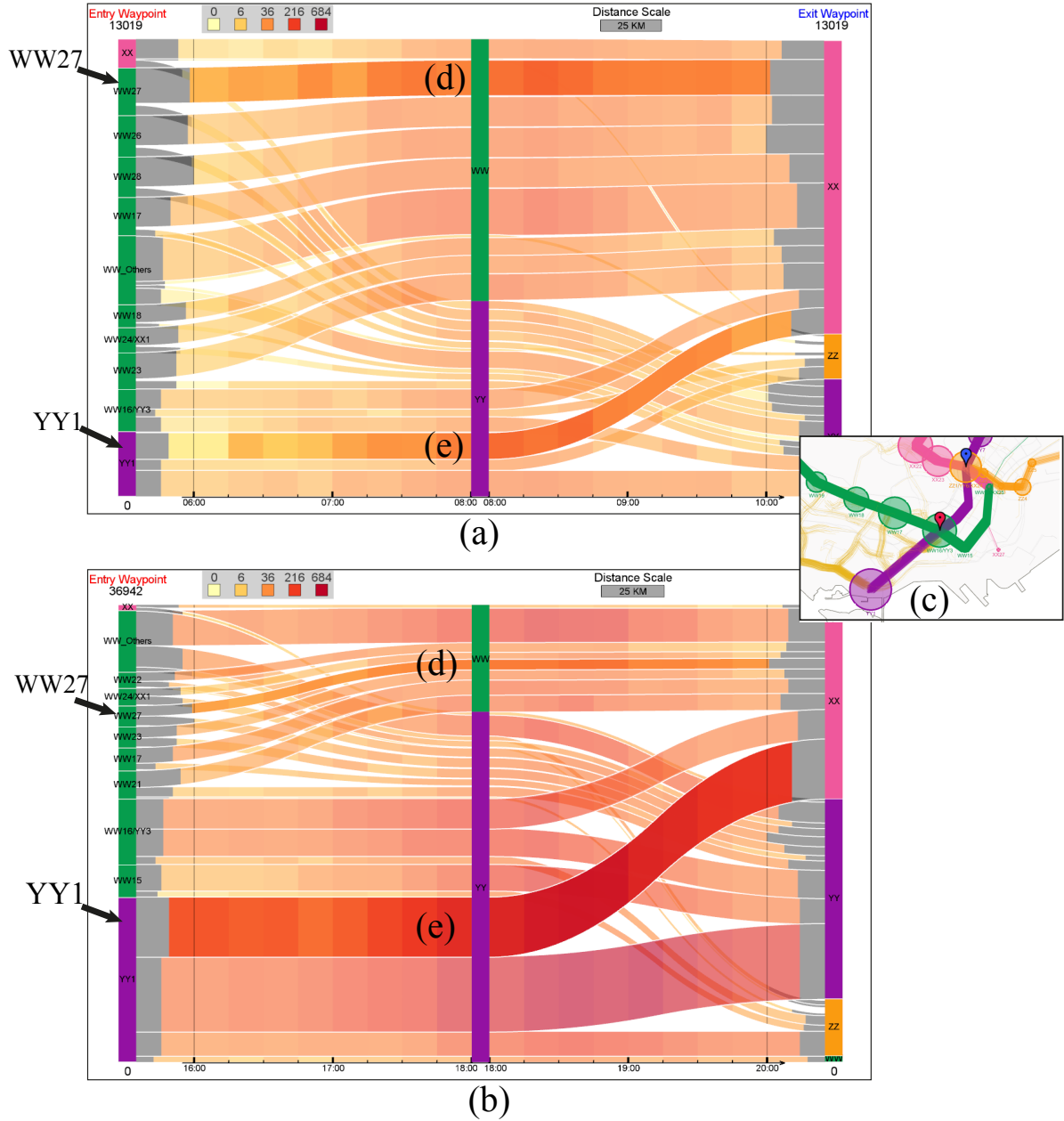


Figure 5.9: Case study 1: transportation network usage analysis. (a) & (b) present the morning and evening OD-flow temporal views, respectively, for the same pair of waypoints shown in the map view (c). Highlighted ribbon (d) shows that trajectories leaving WW27 station for XX-line stations (red) always pass through the WW branch (green) instead of the YY branch (purple) (note: they are two different time-efficient paths between the waypoints, see again the map view), while (e) shows that trajectories transferred from YY1 station to XX stations always pass through the YY branch.

contributes a slightly larger flow volume to stations in the XX line than YY1 in the morning; while from Figure 5.9(b), we can see that the ribbon from YY1 to XX line becomes much wider than that from WW27 to XX line, indicating more people coming from YY1 to XX line in the evening.

Our visualization can also support task *T5-3*. At a glimpse, we can see that Figure 5.9(a) contains less red colors, indicating less flows through the two waypoints in the morning than in the evening. Second, when further examining the two figures, we can see that the colors around 08:00 in Figure 5.9(a) and around 18:30 in Figure 5.9(b) are more dark red than others. This reveals 08:00 as the morning peak time and 18:30 as the evening peak time for flows across the selected waypoints. Moreover, looking at the height of YY1 in Figure 5.9(b) (bottom left), we can see that it has more flows as compared to other origin stations. This may be due to the fact that YY1 is a popular shopping area near Sentosa in Singapore; more people return home from YY1 in the evening.

Path-related information can also be explored in our visualizations for supporting task *T5-4*. From Figure 5.9, we can find that people choose the intermediate paths (WW or YY) mainly depending on their origins and destinations. Taking the flows from WW27 to XX stations as an example, i.e., Figure 5.9(d), almost all passengers chose the WW path instead of the YY path, but they chose the YY path if their destinations in the YY stations. Lastly, Figure 5.9(b) also shows that there are relatively more flows through the YY line than the WW line in the evening, and most of these flows either start from YY1 or end at the stations in the YY line.

5.5.3 Study 2: Daily Pendulum Movements Exploration

In transportation, pendulum movements describe an obligatory urban mobility pattern that is highly predictable and recurring on a regular basis [104]: Employees who commute from residential to working areas contribute to the A.M. peak flow; when they return home, they

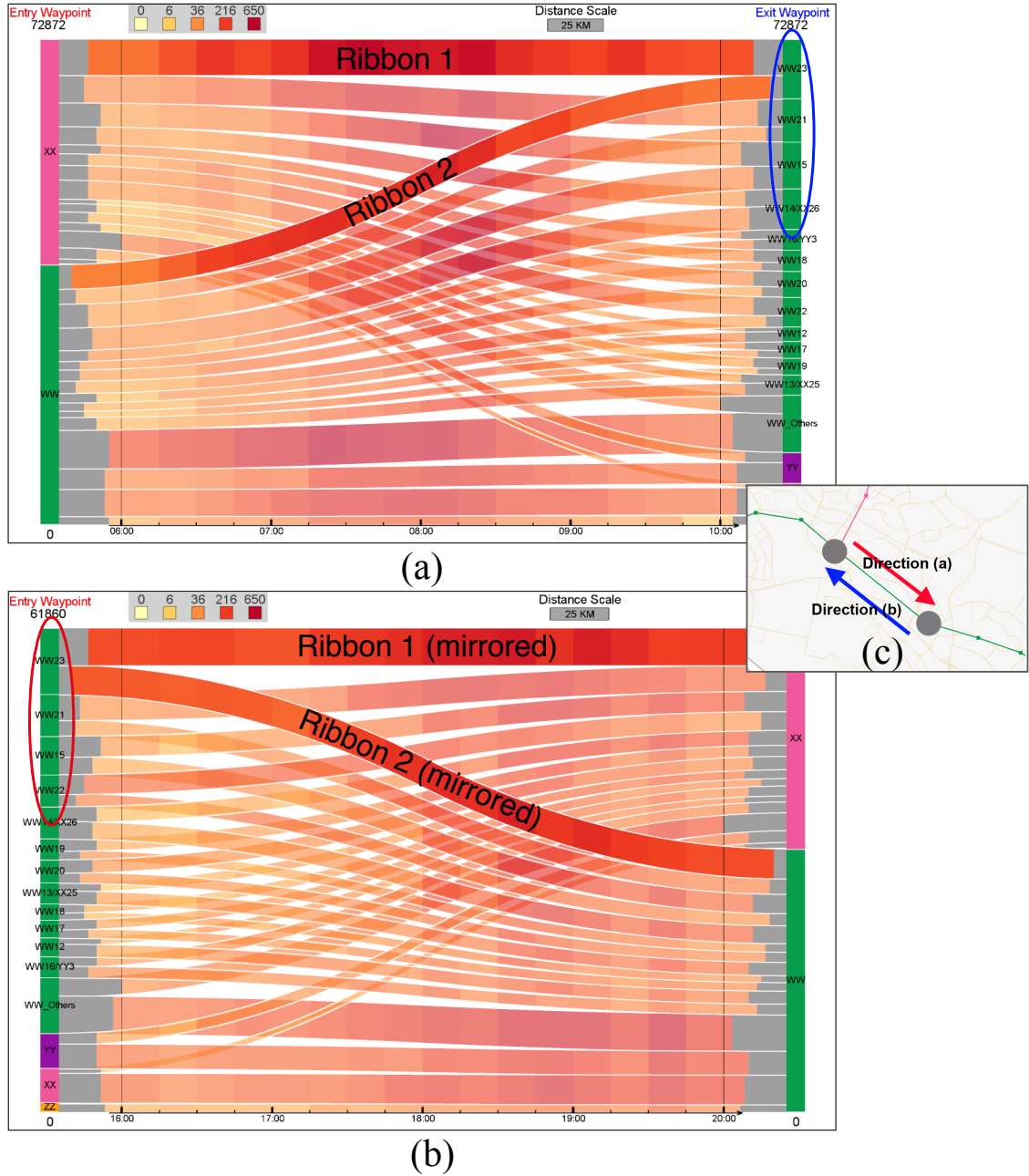


Figure 5.10: Case study 2: daily pendulum movement exploration. (a) & (b) present an interesting pendulum movement pattern, which illustrates home-to-work movements through the red arrow in (c) in the morning (06:00-10:00), and work-to-home movements through the blue arrow in (c) in the evening (16:00-20:00).

contribute to the P.M. peak flow. By analyzing this pattern, researchers can effectively measure the distribution of residential and business regions [84]. Thus, exploring pendulum movements is highly valuable for transportation planning.

The pendulum movement pattern is mainly related to analytical task *T5-1*, since we need to determine the locations of the origins and destinations, and then check whether they swap roles in the morning and evening periods. In this study, we explore the daily pendulum movement patterns in the Singapore MRT data on a normal working day.

Here, we first specify stations WW24/XX1 and WW23 as the entry and exit waypoints (red arrow in Figure 5.10(c)) and produce the waypoints-constrained OD view for the morning period 06:00-10:00 (Figure 5.10(a)). After that, we swap the role of the two waypoints (blue arrow in Figure 5.10(c)) and produce another view for the evening period 16:00-20:00 (Figure 5.10(b)). By comparing Figure 5.10(a & b), we can observe interesting pendulum movement patterns.

First, both views identify nearly the same set of origins and destinations but with swapped roles. In the morning, WW and XX stations generate similar amount of flows, indicating that the areas around both sets of stations have similar residential population. Moreover, most of these flows end at four specific stations, see the blue circle in Figure 5.10(a), indicating that these stations mostly locate in business areas as compared to others. Furthermore, while in the evening, most of the flows from the four stations (see the red circle in Figure 5.10(b)) have similar flow volumes, mirroring the flows in the morning, suggesting that most people follow reversed routes to return home, so these flows mostly end at WW and XX stations with similar flow volumes.

Second, flow volumes between the same OD pairs in A.M. & P.M. with reversed directions are almost the same. Taking the highlighted ribbons connecting XX-line to WW23 (*Ribbon 1* in Figure 5.10(a)) and WW-line to WW23 (*Ribbon 2* in Figure 5.10(a)) as examples, we can find similar flow volumes for each ribbon as compared to its mirrored counterpart in Figure 5.10(b).

This further shows that most employees return home from workplace through reversed routes. It would be interesting to explore whether such a pattern also happen in other big cities such as London and New York.

5.5.4 Expert Interview

We conducted expert interviews with five transportation experts: two senior researchers with 15+ years of research experience (denoted as SR1 & SR2) and three junior researchers with less experience (denoted as R1, R2 & R3). Since this research work is conducted through a transdisciplinary research programme (Future Cities Laboratory), which comprises computer scientists, transportation researchers, architects, etc., the first author can easily reach out to transportation researchers in the institute. Here, SR1 is one of the co-author of this paper while the other experts are independent researchers from the institute.

In the expert interview, we started with a few questions to identify their background and explained our interface design and visual encodings. We then showed the two case studies and asked for their feedbacks. Each interview lasted for 1 to 1.5 hours, and their feedbacks are summarized as follows.

Visual design and interactions. In general, all experts agreed that our visual analytics interface is nicely designed, and supports the analytical tasks well. They especially liked the OD-flow temporal view since it can help to reveal both the OD flows over the whole time period as well as in a specific time interval. Normally, they employ conventional flow maps that connect origins and destinations when studying OD patterns. They pointed out that the conventional flow map could easily cause visual clutter with that many OD pairs and that they have to produce many views to compare the ODs at different time periods. SR2 said “*I never thought one single view can clearly present the OD flows and their temporal variations.*” SR1 specially appreciated the order of OD-pair ribbons with larger volumes in front, as “*in general, OD pairs with larger volumes are more interesting*” to them.

The experts also acknowledged the usefulness of the in- and out-flow views as visual aids for exploring the trajectory paths. R3 commented the views: “*intuitively demonstrate the passengers accumulate and spread along the network.*” SR1 & R3 pointed out that being able to observe the travel time from each origin/destination to the entry/exit waypoint is very useful, as it reveals passenger’s preference regarding travel time. “*There must be something behind if many passengers need to travel long times,*” said SR1.

The experts appreciated the interactions offered by our interface. There can be easily hundreds of OD pairs in OD analysis. Being able to *filter* unimportant and *highlight* important information “*would greatly facilitate my analysis,*” said R2. The experts also liked the aggregate/disaggregate interactions, which can reduce the number of OD pairs, allowing them to explore particular ODs on their demand.

Suggestions. The experts gave some fruitful comments to improve our interface. They mentioned that we can provide more spatial information in the waypoints-constrained OD view, such as a dimmed map on the background. SR1 & SR2 also hoped that our system would support some in-depth analysis of some mobility information. For example, they would like to explore if passenger travel distances follow the power-law distribution [25], yet our visualization can only present relative (average) travel distances among the OD pairs. The experts also had some concerns about adopting our interface to more complex subway systems, which do not come with a simple color coding scheme. Nevertheless, they agreed if we can pre-define subway line colors and do some training, the users would get used to our system.

5.5.5 Discussion

In this work, we explored waypoints-constrained OD patterns of passenger flows in the Singapore MRT network. Case study 1 demonstrated that our visualization can effectively present OD-pair flows, temporal- and path-related information, with respect to analytical tasks T5-2,

T5-3, and *T5-4*, and case study 2 showed that our design can also support well analytical task *T5-1*. The expert feedbacks further commented the effectiveness of our visual analytic system.

We believe that our system can be adopted to visualize OD patterns of movement data in more complex networks: For example, in a general case, we can hierarchically partition the geographical space into regions based on administrations or methods like [50, 10], and explore OD flows in-between these regions. The experts also highlighted that geographical partitioning method could be aligned with their traditional OD analysis.

The OD-flow temporal view depicts large amount of information, yet the design may fail when given excessive quantity of OD pairs. Due to the number of resulting ribbon crossings, exploring OD patterns with more than forty OD pairs altogether (without hierarchical grouping) is not recommended. Nevertheless, our visualization design is suitable for waypoints-constrained OD pattern analysis for two reasons. First, according to previous study [132], the number of origins and destinations that are interested to transportation researchers is limited in most situations. Second, we adopt the “*overview + detail*” principle, allowing users to interactively control and manipulate OD pairs being presented. Most existing OD visualization methods, e.g. Flowstrates [22], do not offer this feasibility.

Not being able to preserve more spatial context can be considered as a limitation of our approach. However, this is a common problem for methods that visualize arbitrary OD flows [7]. To mitigate the spatial information loss, we offer the in-flow and out-flow views to facilitate the exploration of trajectory paths.

Chapter 6

Visual Analytics for PTS Mobility

Due to their increasing complexity of PTS, designing effective methods to visualize and explore PTS is highly challenging. Most existing techniques employ network visualization methods and focus on showing the network topology across stops while ignoring various mobility-related factors such as riding time, transfer time, waiting time, and round-the-clock patterns. The work in this chapter aims to visualize and explore commuter mobility in a PTS with a family of analytical tasks based on inputs from transport researchers. After exploring different design alternatives, we come up with an integrated solution with three visualization modules: *isochrone map view* for geographical information, *isotime flow map view* for effective temporal information comparison and manipulation, and *OD-pair journey view* for detailed visual analysis of mobility factors along routes between specific origin-destination pairs. The isotime flow map linearizes a flow map into a parallel isoline representation, maximizing the visualization of mobility information along the horizontal time axis while presenting clear and smooth pathways from origin to destinations. Moreover, we devise several interactive visual query methods for users to easily explore the dynamics of PTS mobility over space and time. Lastly, we also construct a PTS mobility model from millions of real commuter trajectories, and evaluate our visualization techniques with assorted case studies with the transport researchers.

6.1 Introduction

Studying the efficiency of a PTS is highly beneficial to both individuals as well as to the entire city as a whole. Thanks to recent availability of various forms of public transport data, including the commuter journey data collected via RFID cards, transit schedule data, and transport network data, we now can study and explore the efficiency of a PTS by modeling and integrating these real-world data rather than relying on simulations. By then, we can further design and develop visual analytics methods to explore these data and serve the transport researchers and urban planners. In particular, this chapter focuses on exploring and visualizing commuter mobility in a PTS, e.g., how fast commuters can travel by PTS, which is a highly crucial factor that impacts the overall PTS efficiency.

However, developing visual analytics methods to meet this goal is a highly challenging task due to the following issues:

- First, public transport systems are increasingly complex to meet the population growth, e.g., metropolises like London and New York have 270 underground stations and 460 subway stations, respectively, offering more than a billion commuter trips per year. If we also consider buses and other transport modes, the PTS network would be overly complex for exploration and analysis. This motivates us to study PTS mobility models [99, 67] from the transport research community to analyze routes started from a common origin in a complex network.
- Second, existing works in visualizing public transport networks mostly employ network visualization methods and focus on presenting the network topology across stops. They ignore various mobility-related factors, e.g., riding time, transfer time, waiting time, etc., that affect the PTS efficiency. Hence, novel methods have to be developed to meet the needs of exploring and analyzing these factors.

- Lastly, the mobility-related factors to be explored are not static. They vary dynamically with both time and space, and could also exhibit round-the-clock patterns. Hence, spatio-temporal visualization strategies have to be considered to maximize the visual analytics capability of a method.

To address the above issues, we present in this chapter a visual analytics framework to visualize and explore mobility-related factors in a public transport system with three visualization modules:

- *isochrone map view*, which presents geographical regions accessible from a given starting location within certain duration;
- *isotime flow map view*, a novel strategy that linearizes a flow map in a parallel isotime fashion, enabling visualization and exploration of various mobility-related factors; and
- *OD-pair journey view*, which enables interactive exploration of various mobility-related factors, and their temporal variations, along the origin-destination journeys.

To develop the above visualization modules, we first analyzed the problem with the help of two transport researchers, and identified the related analytical tasks (Section 6.2). Then, we constructed a PTS mobility model from different pieces of real data including transport network data and commuter RFID card data with several million trips (Section 6.3). After that, we developed the three visualization modules mentioned above, and refined their visual designs with the transport researchers (Sections 6.4.1 to 6.4.3). During the design phase, we also implemented and explored different design alternatives for presenting mobility-related factors (Section 6.4.4). Lastly, to evaluate our visual analytics framework, we explored two case studies with transport researchers who are currently working on public transport planning and management, and also presented expert feedbacks received from two transport researchers (Section 6.5).

6.2 Overview

In this section, we first present the problem definition. After that, we describe the related analytical tasks, the mobility-related factors, and the input data set, and then give an overview of the system workflow.

6.2.1 Problem Definition

In land-use and public transport planning, transport researchers would like to explore the level of connections, or the travel efficiency, from a particular location to other parts of the city, given the existing land use and transport network. By this, they can quickly identify which part(s) of the city is/are less developed, find out what facilities are lacking, and explore inefficient usage of public transport resources. This problem is also related to optimal routes algorithms [99, 67] in transport research, where transport researchers study routes starting from a common origin to different points of interests in the city.

This is a collaborative work with two transport researchers specialized in public transport systems. Based on their inputs, the following visualization problem is defined:

- **Input:** an origin A in the given public transport network, starting time t_0 , and a certain time duration T ;
- **Output:** a set of destinations $\mathbf{B}$ (and related travel routes) that are reachable from A at t_0 within T ; and
- **Goal:** we aim to present and explore mobility-related factors (see Section 6.2.2) associated with the travel routes from A to $\mathbf{B}$.

6.2.2 Analytical Tasks and Mobility-Related Factors

To address the problem, here are the basic analytical tasks that our visual analytics interface should support:

- **Task 6-1:** Given the input information A , t_0 and T , extract and present all reachable destinations $\mathbf{B}$;
- **Task 6-2:** Present clear pathways/routes from A to $\mathbf{B}$; and
- **Task 6-3:** Examine and compare the travel time and travel efficiency of the routes from A to $\mathbf{B}$.

The above basic tasks focus on presenting and exploring routes starting from A at a given time t_0 . Additionally, we would need to allow the users to select specific destination nodes, say $B_i \in \mathbf{B}$, and then:

- **Task 6-4:** Present detailed path information from A to B_i , i.e., various mobility-related factors, see below for details; and
- **Task 6-5:** Examine the mobility-related factors and their round-the-clock pattern, i.e., their temporal variations over a day.

In particular, the following mobility-related factors are considered:

- *Waiting time* at a bus stop or subway platform for a route service;
- *Riding time* on a vehicle for traveling between two stops;
- *Transfer time* for walking between neighboring stops; and
- *Travel efficiency* measures the efficiency of traveling between a specific pair of origin and destination relative to the average efficiency (speed) of travel in the PTS.

The above analytical tasks and mobility-related factors are the baseline requirements for our visual design to be presented in Section 6.4.

6.2.3 System Workflow

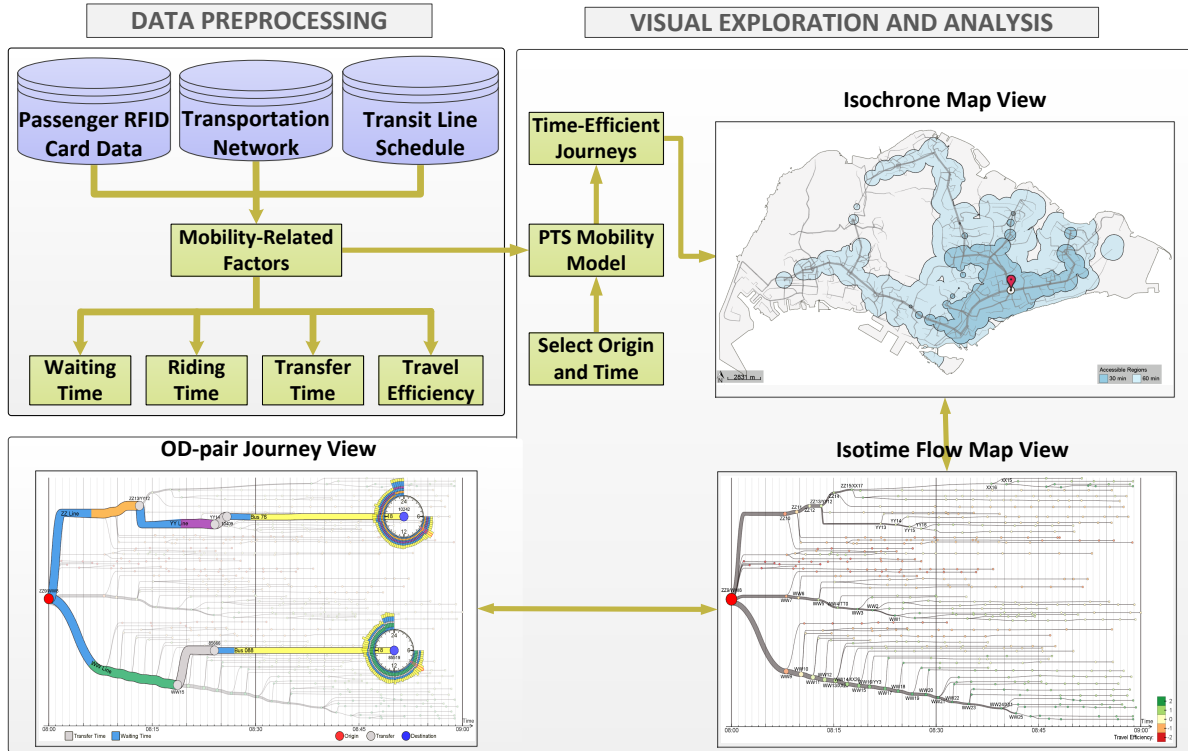


Figure 6.1: Overview of our system workflow. In the data preprocessing stage, we integrate the input data to estimate various mobility-related factors. These information are then presented in the visual exploration and analysis stage, which comprises three visualization modules: (i) isochrone map view, (ii) isotime flow map view, and (iii) OD-pair journey view, which complement one another and work together to support various analytical tasks.

The workflow of our visual analytics framework is illustrated in Figure 6.1. It has two major phases: The data preprocessing phase loads and integrates various input data, estimates mobility-related factors over space and time, and constructs the PTS mobility model, see Section 6.3. To support near real-time determination of routes from a given origin **A** to destinations **B**, we index the mobility-related factors on stops and transit routes. Note that this preprocess-

ing phase is a one-time offline process, after which we store the precomputed information on hard disk, and load them in the next phase.

The visual exploration and analysis phase starts with our main interface with three modules: isochrone map view (Section 6.4.1), isotime flow map view (Section 6.4.2), and OD-pair journey view (Section 6.4.3), which complement one another and work together to present the mobility-related factors and support the various analytical tasks.

6.3 Modeling PTS Mobility

In this section, we first describe the PTS mobility model employed, and then present the data preprocessing stage, which focuses on estimating the mobility-related factors from the input data.

6.3.1 PTS Mobility Model

In reality, public transport stops are usually not the origin or destination of a commuter journey; one often needs an initial walk, say from home/office/shop to a public transport stop, before the PTS trips and transfers, as well as a final walk to reach the destination. Since we have no data about the initial and final walks, we consider commuter journeys to start and end at stops in the public transport network.

One key factor that affects how commuters plan their journeys is travel time, which is also a crucial factor that affects the overall efficiency of a public transport system. Hence, we choose to construct a PTS model [99, 67] that focuses on time-efficient journeys. In detail, we model a commuter journey with n trips and $n-1$ transfers, so the overall travel time of the journey is modeled as

$$T_{journey} = \sum_{i=1}^n T_{trip}^i + \sum_{i=1}^{n-1} T_{trans}^i, \quad n \geq 1,$$

where T_{trip}^i is the travel time for the i th trip, and T_{trans}^i is the transfer time between the i th and $(i+1)$ th trips. Since waiting time is often needed before boarding a vehicle, e.g., train and bus, we further divide T_{trip}^i into waiting time T_{wait}^i and riding time T_{ride}^i :

$$T_{journey} = \sum_{i=1}^n [T_{wait}^i + T_{ride}^i] + \sum_{i=1}^{n-1} T_{trans}^i, \quad n \geq 1.$$

6.3.2 Estimating Mobility-Related Factors

In the data preprocessing stage, we first clean the raw commuter trip data by removing incomplete and erroneous data records, e.g., some commuters went out of buses without tap-out, some commuters stayed exceptional long inside the metro system compared to normal travel time needed to go between their tap-in and tap-out stations, etc.

Since mobility-related factors are time-varying, it is not feasible to estimate their continuous changes over time even with millions of commuter trip records. Hence, we divide the temporal dimension into seventy-two 15-minute time bins from 6am to midnight, which is the normal PTS operating period of a day. Then, we integrate various input data, and estimate the average value of each mobility-related factor (per stop or stop connection) per time bin. Note also that since commuter tap-in and tap-out mechanisms are slightly different for buses and metro services, we may need to consider bus and metro independently when estimating the mobility-related factors. Moreover, we assume that metro services always follow the transit line schedule while bus services may not (due to road sharing and local traffic).

Waiting Time is a per transit route, per stop and per time bin quantity. To estimate it for bus services, we extract all bus trips from the RFID card data. For each stop of each bus transit

route, we first compute the average time over all tap-in tap-out times at the same stop of the same bus to estimate when the bus stays at each stop, say t_{bus}^i . Hence, we can obtain all t_{bus}^i for all bus services (same transit route) at a given stop over the day, and then compute the time interval between successive t_{bus}^i to estimate the bus frequency (interval) at each stop per bus line per time bin; half of such a value is the expected waiting time.

For MRT services, though transit line schedule data is accurate, actual waiting time may sometimes be longer than the time interval between successive trains since during the peak hours, commuters may not be able to board a train immediately after reaching the MRT platform. Hence, we estimate MRT waiting time as follows. First, we extract all MRT trips without MRT-to-MRT transfer since having a transfer could bias the computation below. Then, for each trip, we extract the tap-in time, and search for the next train that the commuter should have boarded at the tap-in station. By this, we can look up the transit line schedule to obtain the riding time required for him/her to reach the destination station, and estimate the related waiting time as: $(\text{tap-out time} - \text{tap-in time}) - \text{scheduled riding time}$. Since we can obtain multiple waiting time from different commuter records, we further compute their average as the expected waiting time.

Riding Time is a per successive stops (along the service route) and per time bin quantity. For bus services, after we estimate t_{bus}^i at the stops of each bus line (see above), we can estimate the riding time of each bus between successive stops of the same bus. Again, we average multiple instances of such a value to obtain the expected riding time per successive stops and per time bin. For MRT services, we obtain riding time simply by looking up the transit line schedule.

Transfer Time. There are three cases of transfer: First, it is from MRT to MRT. If the transfer happens between nearby platforms, we assume zero transfer time. However, in some cases, one may have to walk a fairly long distance from one platform to another. Since there are no card tapping activities during the transfer, we estimate transfer time by taking advantage of the

data massive-ness: 1) extract all MRT journeys with only one transfer; 2) estimate the transfer time of the journey as: $(\text{tap-out time} - \text{tap-in time}) - (T_{wait}^1 + T_{ride}^1 + T_{ride}^2)$; and 3) again, average the results from different journeys per time bin.

The second case is from bus/MRT to bus. If the two bus stops are the same (same reference ID), we assume zero transfer time. Otherwise, we need a walk to the next bus stop, so we estimate the transfer time as $\text{tap-in time (next bus)} - \text{tap-out time (prev. bus)} - \text{wait time (next bus)}$. Note that MRT to bus is slightly different from bus to bus since it requires a walk from MRT platform to the tap-out gate. However, since we have no information about such a walk, we ignore it and estimate transfer time in the same manner as bus to bus.

The last case is from bus to MRT, where we estimate the transfer time as $\text{tap-in time (MRT)} - \text{tap-out time (bus)}$.

Travel Efficiency differs from the general concept of *speed* since it considers also waiting and transfer time in addition to riding time. Moreover, it describes the relative efficiency of travel along a specific route as compared to the mean mobility of the entire PTS network.

Before computing the travel efficiency of a specific route, we first determine the mean mobility of the entire PTS by: 1) compute the mobility of each commuter journey in the RFID data as: total journey distance divided by $T_{journey}$; and 2) compute the mean mobility μ and also its standard deviation σ over all the journeys. By this, the *travel efficiency* of a given route (started at a given time) is obtained by normalizing its mobility value against μ and σ .

6.4 Visualization Design

In this section, we describe how we support the analytical tasks defined in Section 6.2.2 through the following three visualization modules:

6.4.1 Isochrone Map View

For Task 6-1, our goal is to extract and present all reachable locations $\mathbf{B}$ given A , t_0 , and T . To handle it, we first compute time-efficient journeys from A to every single stop in the PTS using the estimated mobility-related factors. This is done by a real-time breadth-first-like mechanism (single-source shortest time-efficient paths) that iteratively identifies and expands time-reachable stops (nodes in the network graph) over the geographical map before T is reached.



Figure 6.2: The isochrone map view presents spatial-reachability regions from 08:00 at origin A (red icon), which is station XX16/WW3 in the city center, by contour lines and areas over the geographical map. Dark blue and light blue indicate $[0, 30]$ and $(30, 60]$ minutes, respectively.

In addition, at every reachable stop (including A), we consider “commuter walk” from the stop by using the remaining journey time at the stop within T , and assume a constant walking speed of 5km/h without encountering obstacles like buildings and roads. Hence, every stop will be surrounded by a circular region; we further union all these regions to determine time-reachable regions on the map (note: such union is done by rendering without tedious geometric computation). After that, we can plot the related contour lines and areas, see Figure 6.2 for an example with a city center location as A . Here we use a red icon on the map to show the

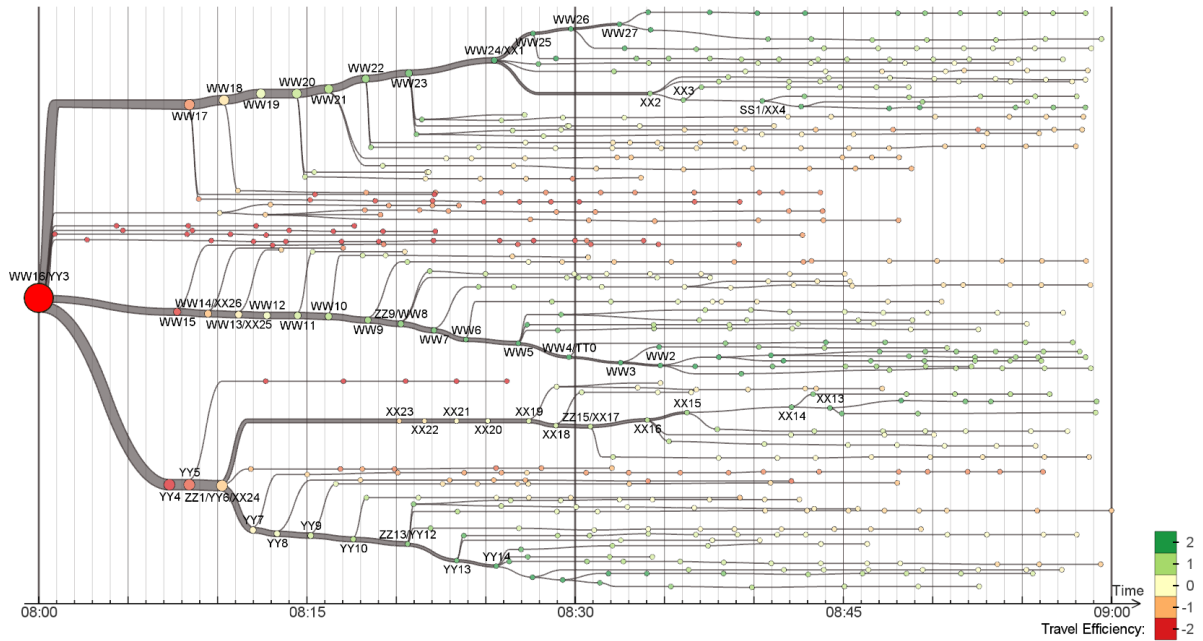
location of A , and highlight the contour regions in blue: dark blue for $[0, 30]$ min., light blue for $(30, 60]$ min., and white for >60 min. Moreover, we present in gray the set of all reachable edges, which is a subset of the entire PTS network, and adjust their line width to reveal the amount of time-efficient journeys that pass through each edge. By this, main branches can be emphasized.

6.4.2 Isotime Flow Map View

To handle Tasks 6-2 & 6-3, i.e., to present clear routes from A to B and to examine and compare their travel efficiencies, the isochrone map view alone is insufficient. If we apply colors to B in this view to show the travel efficiency, the colors we employed would easily mess up with the isochrone colors. Moreover, it is hard to present clear pathways for examining and comparing time-efficient journeys in the isochrone map view, particularly with numerous pathways from A to B .

Hence, handling Tasks 6-2 & 6-3 is non-trivial, so we first explore different design alternatives in a pilot study, see Section 6.4.4 for detail. After comparing these alternatives, we propose the *isotime flow map view*, which is a novel visualization strategy, that presents a flow map visualization in a parallel isotime fashion, see Figure 6.3 for an example. The followings detail its construction procedure:

- (i) **Parallel isotime model.** First, we arrange a horizontal timeline on the bottom of the view to show the journey time from t_0 and $t_0 + T$ (left to right). In this view, A is the red dot on the left while all destinations and nodes to B (in fact, all visual elements) are tagged with time. Thus, we can quickly look at the horizontal coordinate of any visual element w.r.t. the time axis to find out the related journey time from A . To further facilitate such visual examination and comparison, we draw an array of vertical gray lines in the background of the view to show different amount of time intervals.



(ii) **Tree structure.** To present time-efficient journeys from A to $\mathbf{B}$, rather than showing them one by one, we present them as a tree structure, which is a subgraph of the entire PTS *rooted* at A . Such tree is constructed by examining the journeys and identifying *branch nodes* (transfer stops) and *leaf nodes* among the journeys. Here we also count the number of time-efficient journeys (as a weight factor) that go through each branch node.

(iii) **Spatial layout.** To present the time-efficiency journeys as a tree structure, we take the flow map visualization approach and layout the tree according to the parallel isotime model. Hence, the horizontal coordinates of all the nodes in the tree are fixed according to the related journey time from A , see Figure 6.3, and so, our main task in this step is to determine the vertical coordinates of all the nodes in the tree. As for this, we devise the following recursive method, which helps to avoid visual clutter and promote tree

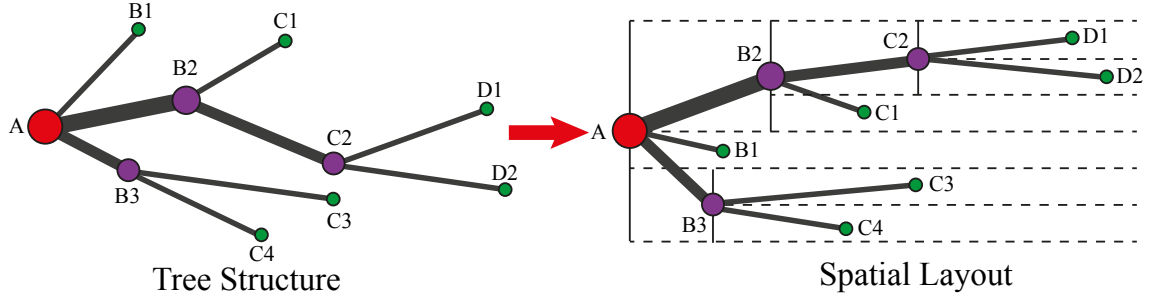


Figure 6.4: Left: A tree structure that includes all time-efficient journeys from starting node A. Right: the result of applying our spatial layout algorithm to arrange the nodes in the visualization view.

balancing:

- Given A, we first extract all child branch nodes of A.
- To improve the tree balancing, we sort these branch nodes as follows: Given k nodes, we first find out the node with the highest weight factor and assign it as n_1 ; the node with the 2nd highest weight as n_k ; the node with the 3rd highest weight as n_2 ; then as n_{k-1} , etc. Let w_i be the corresponding weight of node n_i .
- Then, we divide the vertical range from A into sub-ranges according to w_i . See Figure 6.4(right): node B_2 with the highest weight on the top, B_3 with the 2nd highest weight on the bottom, etc.
- Lastly, we repeat the above procedure for each child branch nodes of A till reaching the leaf nodes.

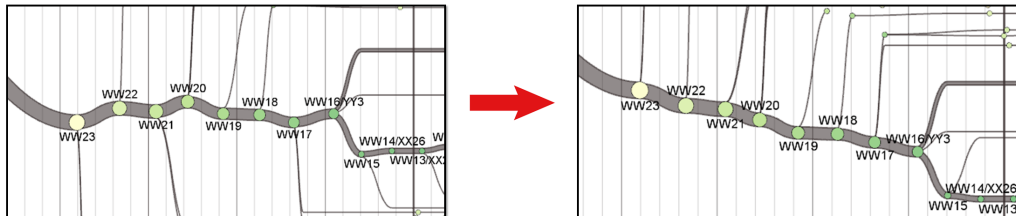


Figure 6.5: Left: up-and-down wobbling. Right: we fix it by vertically shifting the nodes by node-based moving-window averaging.

Note also that to reduce up-and-down wobbling along consecutive branches, see Figure 6.5, we use simple node-based moving-window averaging to slightly shift the vertical position of the nodes.

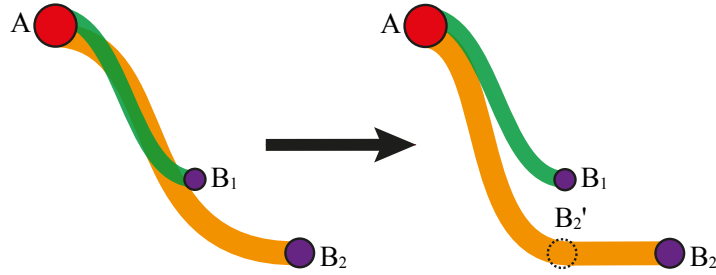


Figure 6.6: Branch routing. Left: branches overlap issue may occur. Right: we resolve it by horizontally shifting the Bézier control points.

- (iv) **Branch routing.** After positioning all the nodes in the view, we next construct a Bézier curve to connect the nodes to form clear and smooth pathways. However, neighboring branches may overlap, see Figure 6.6(a). To resolve this issue, we examine Bézier curves among sibling branches; if an overlap occurs, we horizontally shift the related Bézier control points, see Figure 6.6(b).
- (v) **Node color and label.** Lastly, we color code each node according to its travel efficiency (see Section 6.3.2), and put text labels (name or stop reference ID) at nodes whose on-screen radius is larger than 5 pixels in a up-down-up-down-etc. along consecutive branches.

6.4.3 OD-pair Journey View

Tasks 6-4 & 6-5 focus on supporting visual analysis of mobility-related factors along specific routes from A. Clearly, if we color-code all flow lines in the isotime flow map view and present also their temporal variation, visual clutter would likely happen. Hence, we allow the users to click-and-select destination node(s) in the isotime flow map view, and then perform Tasks 6-4

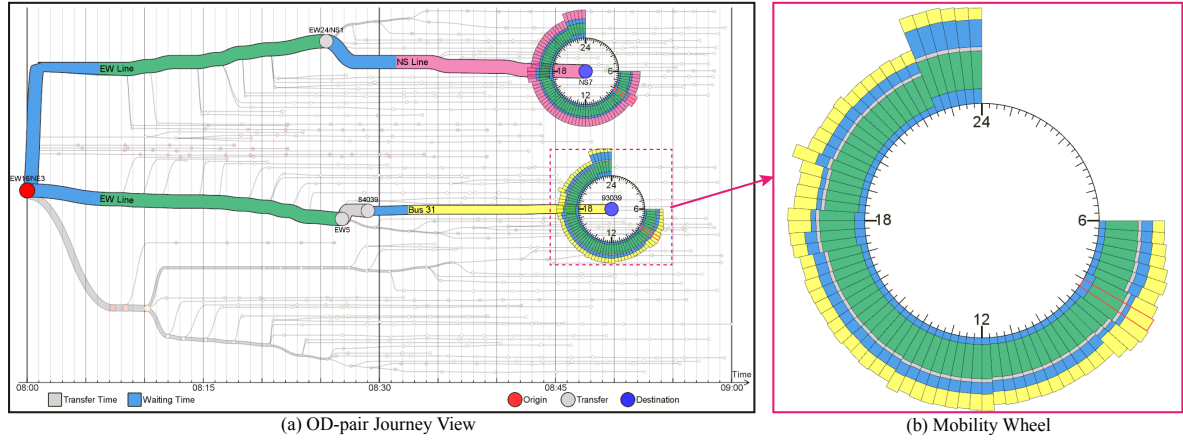


Figure 6.7: The OD-pair journey view focuses on Tasks 6-4 & 6-5, presenting detailed mobility-related information along routes from A to user-specified nodes (left), as well as their round-the-clock variations using our proposed visual representation: *mobility wheel* (right). Note also that we employ the standard colors of Singapore MRT lines to show the riding time on MRT, and encode bus lines by yellow.

& 6-5 through the OD-pair journey view, which is an overlay on the isotime flow map view, see Figure 6.7:

- To support Task 6-4, we need to present detailed mobility-related information in the parallel isotime flow map: 1) we widen to highlight the branches along the user-selected route(s); 2) we color-code different portions of the flow line(s) to show the related mobility-related conditions: light blue for waiting, gray for transfer, standard colors of Singapore MRT lines for MRT riding (e.g., green for XX line), and yellow for bus riding; and 3) we highlight the starting, transfer, and ending nodes in red, gray, and dark blue, respectively, and label them with corresponding reference IDs/names.

Figure 6.7(a) shows two user-selected routes: both routes have similar initial waiting time in XX16/ZZ3 station, but have different waiting times at their transfer nodes. For the journey to NS7 station on top, though it has no transfer time (nearby MRT platforms), it has much longer waiting time than the other route.

- To support Task 6-5, we design a visual cue called the *mobility wheel*, which is inspired

by [83, 123], to show round-the-clock temporal variation of mobility-related factors, see Figure 6.7(b). Our key idea here is to stack the mobility-related factors as small vertical bars, and then pack them in a time bin by time bin fashion around the mobility wheel. By this, we can visualize round-the-clock variation of all contributing mobility-related factors altogether.

In detail, we put a mobility wheel at each user-selected destination when the OD-pair journey view is brought up, see again Figure 6.7(b). In addition, we use the same color coding scheme for showing the mobility-related factors as in Task 6-4, and highlight the current time bin (according to the main visualization) in the wheel by a thin red rectangle. The radius of the mobility wheels remain 200×200 on screen, so the user can zoom in and separate out overlapping wheels if wheel overlap occurs; moreover, the user may also click on a wheel to bring it to the top layer.

6.4.4 Design Alternatives

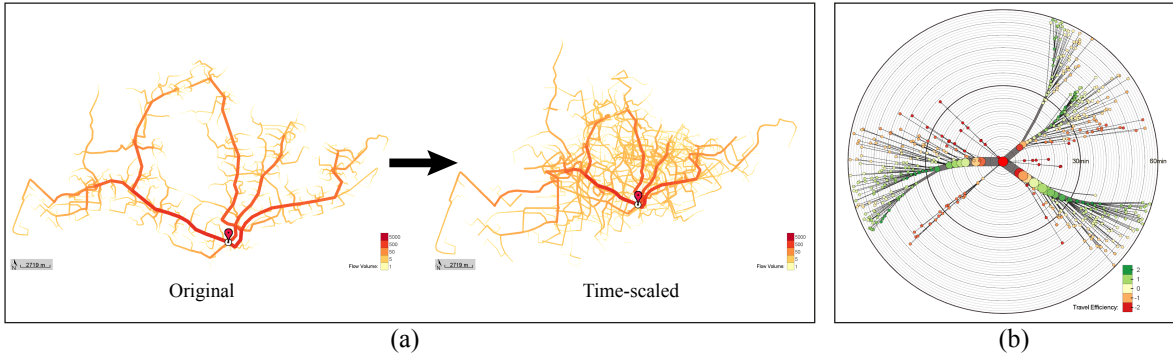


Figure 6.8: Designs alternative that we have explored for handling Tasks 6-2 & 6-3: (a) time-scaled network distortion and (b) radial isotime layout.

As discussed earlier in Section 6.4.2, when we design the isotime flow map view to handle Tasks 6-2 & 6-3, we explore different design alternatives to present the flow map and time-efficiency journeys from **A** to **B**. This section presents how we devise and implement these

alternative designs, and compare them with the parallel isotime layout we have chosen. Below are the two design alternatives we explored:

- (i) **Time-scaled network distortion** deforms the 2D map, so that distances between points on the deformed map relate to travel time [1], see Figure 6.8(a). This method was popularly used in transport to show travel time between locations, and some methods [112, 16, 113] have been proposed to perform the deformation.

In this work, we develop the time-scaled network distortion by a breadth-first visit from *A*: immediate child nodes of *A* are shifted to reflect their travel time from *A*, and we recursively repeat this for the branch nodes until the leaf nodes in the tree structure.

The first problem of this approach concerns with the complexity of the PTS network. When we consider many time-efficient journeys from *A* in the transport network, severe visual clutter would easily occur. Hence, this approach cannot present clear pathways from *A* to the reachable stops **B** (against Task 6-2). Second, it is difficult to accurately compare the travel efficiency of different routes in the visualization (against Task 6-3) even though we know that the total length of the (zigzag) routes relate to travel efficiency, see again Figure 6.8(a). To the best of our knowledge, none of the existing time-scaled method can handle these two issues.

- (ii) **Radial isotime layout.** Besides the parallel isotime layout, we implemented and explored another layout alternative: a radial layout, see Figure 6.8(b). It can be constructed in a way similar to the parallel isotime layout, but it positions *A* at the center, and **B** on concentric circles with increasing travel time along radial direction from *A*.

Comparing with the parallel isotime layout, radial layout can still present clear pathways from *A* to the reachable stops, but after we show this early design to the transport researchers, several negative feedbacks were received from them: First, concerning Task 6-3, it is not intuitive for them to examine and compare the travel time as routes

and mobility-related information are arranged in a radial fashion. According to Heer et al. [57], encoding time progress from left to right along the horizontal axis can aid comparison of time-series events and their trends. Second, the radial layout cannot make full use of the screen space as the aspect ratio of common displays, e.g., 16:9. Lastly, in the parallel isotime layout, we can drag the horizontal time axis to left/right to intuitively modify the current time of the visualization, i.e., t_0 , but for radial isotime layout, such an operation is not intuitive.

6.5 Evaluation

This section presents 1) two case studies on exploring PTS mobility with our interface and 2) feedback from transport researchers.

6.5.1 Case Study 6-1: Spatial Variation of PTS Mobility

In land-use and transport planning, researchers and urban planners are interested in exploring the travel efficiency from a selected location to other parts of the city, as in Tasks 6-1 & 6-3. By the visual analysis, they can know what is lacking and also how to improve.

In this case study, we demonstrate how our system facilitates the exploration of PTS mobility over different locations in a city. Figure 6.9 and Figure 6.10 present the isochrone map views and isotime flow map views related to two different locations on the map: (a) an MRT station and (b) a rural-area bus stop. Though the starting time for both visualizations is 8am, the isochrone map views reveal very different sizes of reachable dominions from the two locations. In fact, the starting location for Figure 6.9 is an interchange MRT station, which is the (XX24/YY1) station in Singapore, with two MRT service lines and a bus terminal nearby. With rapid MRT services, commuters can reach very long distances from this location within a short

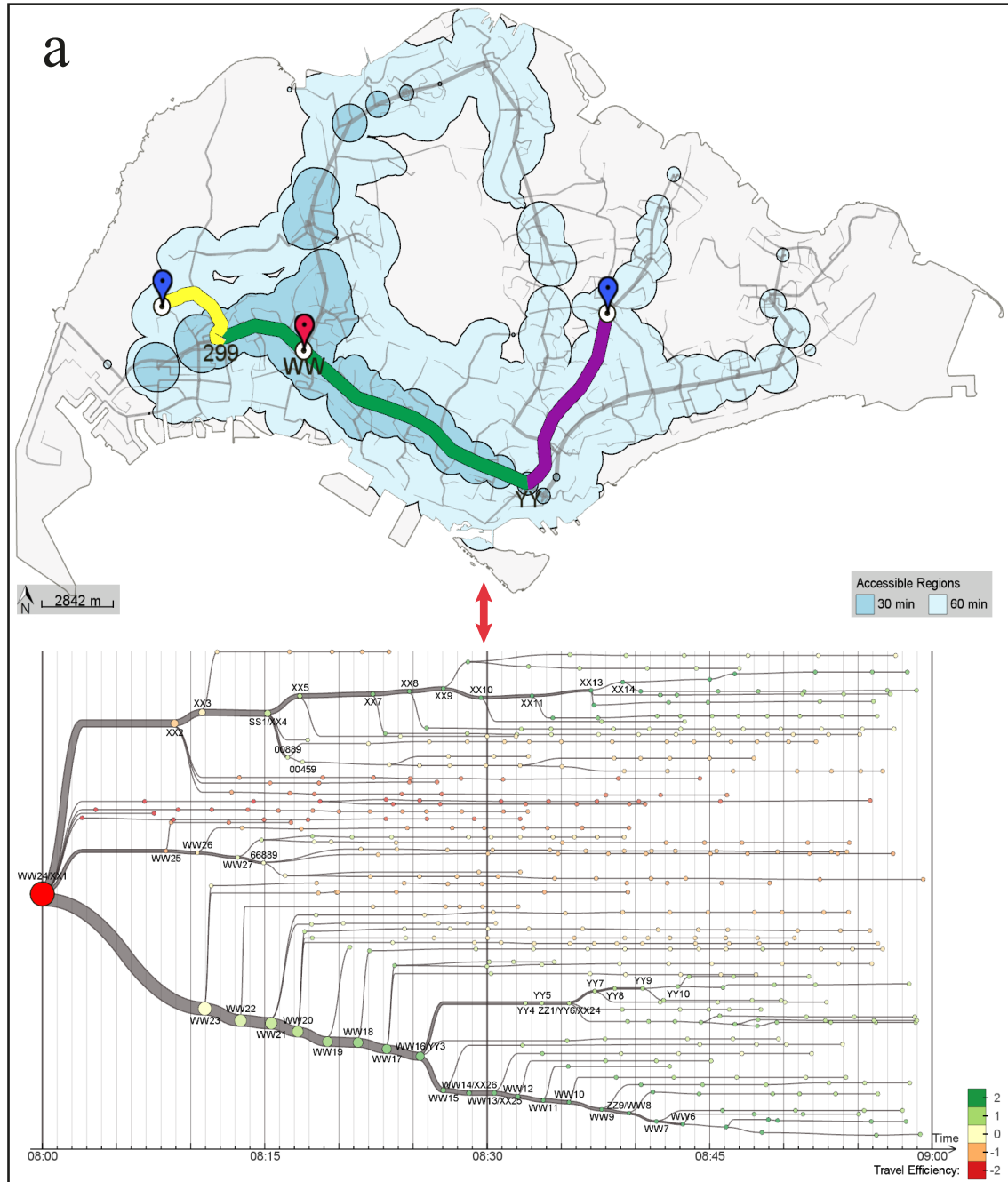


Figure 6.9: Case Study 6-1(a): Exploring the spatial variation of PTS mobility from *MRT* station “XX24/YY1” at 08:00 in the morning.

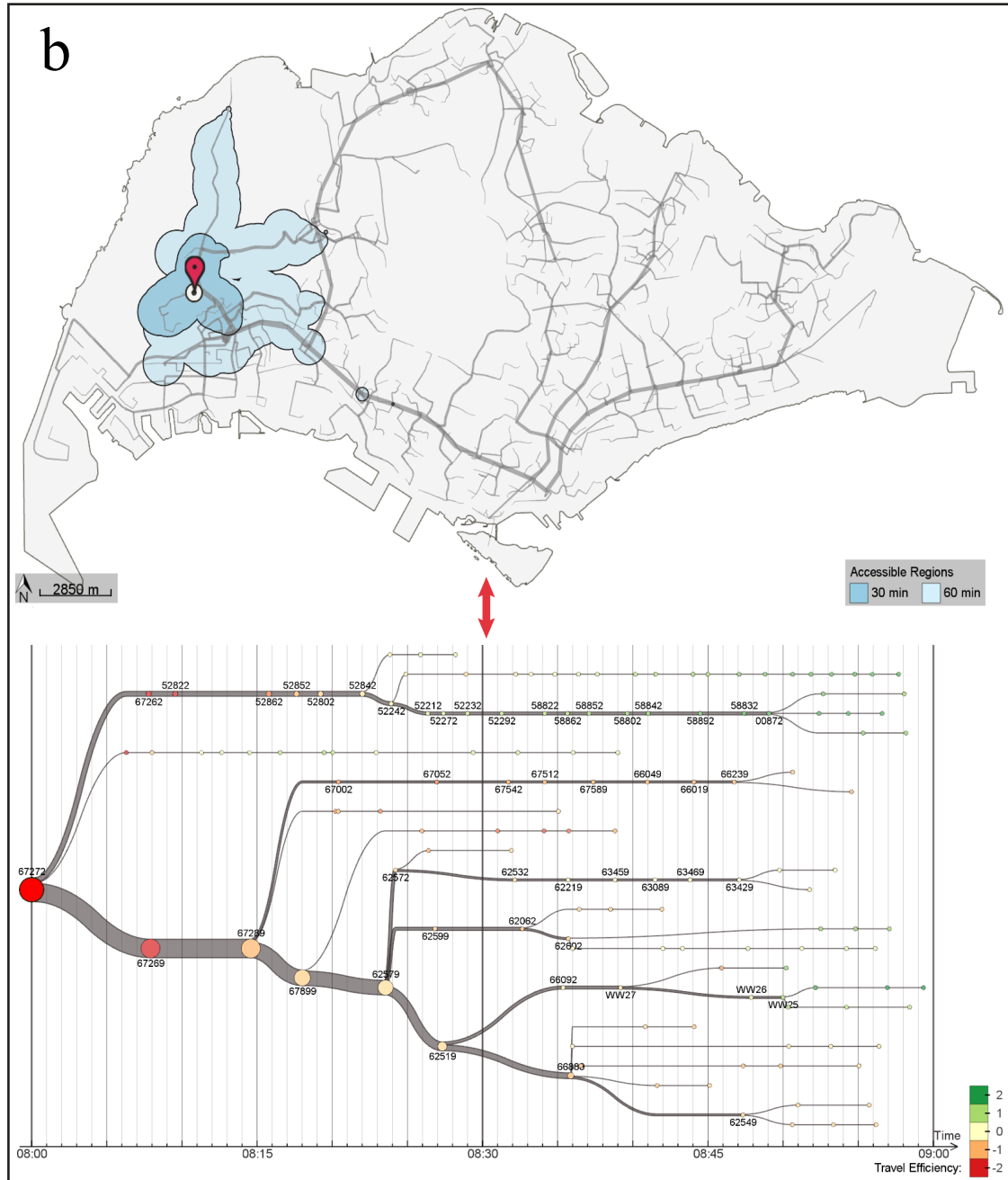


Figure 6.10: Case Study 6-1(b): Exploring the spatial variation of PTS mobility from a rural-area bus stop at 08:00 in the morning.

period of time. In contrast, the starting location for Figure 6.10 is a rural-area bus stop with only two bus lines available. From the isochrone map view on the right, we can clearly see the reachable regions along the two major directions from the bus stop corresponding to the two available buses. Moreover, the bus service line towards the south-east direction relates to more time-efficient journeys since its line width is wider than the other direction from that bus stop. Besides, the node colors in the isotime flow map view also confirm that the local travel efficiency here is very low, as compared to that in Figure 6.9.

6.5.2 Case Study 6-2: Analyzing Mobility-Related Factors

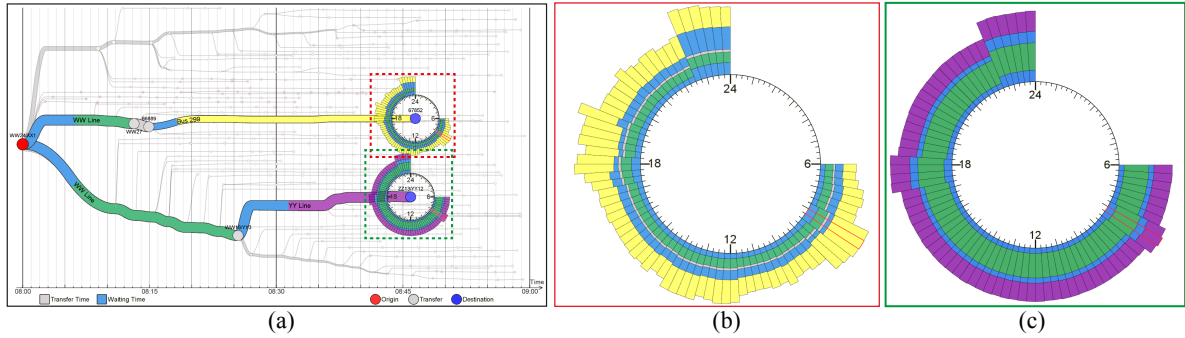


Figure 6.11: Case Study 6-2: Analyzing detailed mobility-related factors along two different user-specified journeys starting from *MRT station “XX24/YY1”* to destinations *Bus stop 67852* and *MRT station “WW13/ZZ12.”* (a) Estimated waiting, riding and transfer time can be clearly presented along the pathways; we can see together the related MRT and bus trips, as well as the transfer points. (b & c) show the zoomed views of the mobility wheels in (a) for presenting the round-the-clock mobility patterns.

Our interface can also allow users to analyze and compare various mobility-related factors that affect the PTS efficiency, as in Tasks 6-4 & 6-5. Figure 6.11(a) presents two user-selected routes from the source location *MRT station XX24/YY1* to destinations *Bus stop 67852* and *MRT station WW13/ZZ12*: the red and blue icons on the isochrone map in Figure 6.9 show their geographical locations, respectively.

From the isochrone map view, we can see that the physical travel distance from *XX24/YY1* to *WW13/ZZ12* is much longer than *XX24/YY1* to *Bus stop 67852*. However, Figure 6.11(a)

clearly shows that their travel time are almost the same with a difference of just a few minutes. To understand how this happens, we can refer to the detailed mobility information shown along the two selected routes in Figure 6.11(a). Here we can find that the main overhead of traveling to *Bus stop 67852* is the riding time spent in the bus trip (yellow segment in the figure) since the bus is relatively slow.

We can further explore the temporal variation of the mobility-related factors using the mobility wheels shown in Figure 6.11(a), see also their zoomed views in Figure 6.11(b & c). In Figure 6.11 (b), we can find huge variations of travel time over a day, with three peaks that correspond to the morning peak hour, evening peak hour, and late night period. The interesting observation here is that riding time in late night period does not vary too much, similar to that of the non-peak hours, but the related waiting time suddenly increases. This finding is also confirmed by the local transport agency since both bus and train service frequencies are halved after 22:30.

On the other hand, the travel time from *XX24/YY1* to *WW13/ZZ12* shows much less variations compared to the other route. Figure 6.11(c) also reveals that the MRT services (green and purple) are relatively more stable over time as compared to the bus services (yellow) shown in Figure 6.11(b). Concerning the peak hours in the morning and evening, the local transport agency told us that the MRT service frequency is doubled during these periods, so waiting time could be reduced. However, the visualizations here show that the waiting time during these peak hours are actually longer than that of the normal periods because commuters may not be able to board a train immediately after reaching the platform during peak periods.

6.5.3 Experts Interview

We interviewed two transport researchers who specialized in public transport systems, and obtained their feedbacks of our interface. One of them is from university E (Expert A), while

another from a Singapore PTS agency (Expert B). In detail, we first explained to them our system workflow and visual encoding, and then demonstrated to them the two case studies we presented above. They both thought that our interface can be a useful tool for planners and operational managers. Their feedbacks are summarized as follows.

Interactive Visual Design. Both of them were impressed by the visual design, especially the isotime flow map view. Expert B commented “It is an excellent idea to display the information of a public transport system in multiple presentation formats under an integrated and interactive manner. Differing the visualization based on the nature of the information item of major concern would greatly enhance the users’ understanding.” Expert A added “the isotime flow map view makes it very easy to identify the time-shortest routes to all destinations that can be reached within a certain travel time threshold, and parallel isotime model makes it easy for him to compare the travel time and efficiency.”

Both experts appreciated the idea of exploring the PTS mobility. Expert A specifically recognized that the choices to select destinations and visualize the detailed mobility-related factors are very useful. He is particularly interested in visualizing transfer information in our interface, as they are strongly negatively perceived travel elements. Expert B pointed out that the ability to switch between different views can greatly enhance users’ understanding of the major information.

Improvements. Both reviewers gave several fruitful comments to improve our system, including providing more visual encoding options for users to select and adding more icons for users to recognize the nodes. Expert A suggested that it might make sense to indicate level of commuter capacity or actual commuter numbers along pathways, instead of showing the number of time-efficient journeys currently in our isotime flow map view. He also recommended us to “show icons or pictograms indicating nearby landmark buildings next to the stops in the isotime flow map view to help user localize the various branches and identify which facilities

can be reached.” Expert B further pointed out that the multiple presentation formats could be explored to emphasize many transport factors, like the boarding and alighting patterns at stops.

Applicability. Expert A also commented: “the tool is highly suitable to support location choice decision processes such as the choice of residence or place for setting up a business.” As shown in *Case Study 1*, the system allows users to evaluate different location options by comparing the ease of traveling with public transport from one to many places. If land use data is also included, it could be extended to be a very powerful tool for site selection in real estate industry. An application example might be that the user would pre-select in what type of places of interest she/he is interested in and the tool would generate an overview of how easy it is to get to various location options.

6.6 Discussion

The case studies demonstrate the applicability of our interface in showing mobility of a PTS. Our current model extracts mobility-related information from massive amount of commuter RFID card data, enabling transport researchers to analyze the efficiency of a PTS based on real data rather than simulations. However, PTS efficiency is affected by many dynamic factors that pose difficulties for transport researchers to recognize and compare. Hence, our interactive method presents various mobility-related factors in an intuitive visualization, allowing researchers to evaluate and compare travel efficiency, as well as to analyze round-the-clock variations and patterns of these factors. Moreover, transport experts can explore the relative travel efficiency of a PTS, and apply the results to land-use and transport planning. Furthermore, such results can also help commuters to make better travel planning through the PTS.

Limitations. First, our current approach assumes no train interruption, so that we can use the MRT transit line schedule data to estimate the mobility factors. Second, we ignore the initial

and final walks taken by the commuters to and from the PTS stops in estimating the mobility factors. Third, our method lacks global perspectives on the transport data, e.g., congestion patterns that affect certain areas of the PTS at certain times, which is an interesting aggregation condition to be explored. Fourth, our current method focuses on one source to many destinations (one-to-many) rather than many sources to many destinations (many-to-many), which is in fact a very challenging problem. Thanks for the reviewer comment that suggests a global overview to browse through possible origins, we will explore and study about it. Lastly, currently we only consider how fast people move, and ignore other factors that affect people's choice, e.g., comfort-ness, cost, etc.

Chapter 7

Conclusion and Future Work

This chapter concludes the thesis, followed by a brief summary of the achievements made and thoughts for future work.

7.1 Conclusion

This thesis has been devoted to developing intuitive and informative visual analytics to depict the knowledge emerged from the input massive urban public transport data. The ultimate goal is to facilitate transport researchers' exploration and analysis of the data, such that to help them manage the traffic flow and improve the PTS design. In particular, three visual analytics systems have been developed, and they are presented in each of the following chapters:

Chapter 4: Visual Analytics for Interchange Patterns

This chapter presents a novel method of visualizing and exploring interchange patterns extracted from the input dataset. First, we present a formal definition of interchange pattern, which can be described as an interchange matrix that summarizes flow volumes of different possible routes across a junction node. After that, we derive from the *circos* figure a new visual representation, namely *interchange circos diagram*, to present interchange patterns. Several practical issues to reduce visual cluttering and to improve the visual analytic capability have

been considered to formulate this design, e.g., bundling bidirectional ribbons and designing statistics boxes to summarize flow volumes. Further than that, we also enhance the visual connection between neighboring diagrams and develop a working interface to present multiple interchange circos diagrams supported with a family of interactions. Lastly, we present two case studies to discuss how our interface can be used to study interchange patterns in the MRT system, and to examine intersection capacity utilization at junction nodes.

Chapter 5: Visual Analytics for Waypoints-Constrained OD Patterns

This chapter studies a local aspect of OD patterns, i.e., waypoints-constrained OD patterns that associates with the trajectories passing through a specific path in a public transport network. The problem is motivated by real-world practical needs, e.g., traffic congestions happen only in specific roads [132]. We develop a novel visual analytics approach, namely *waypoints-constrained OD visual analytics*, to explore and analyze the patterns. First, we model the problem with a pair of user-specified entry and exit waypoints, which can be interactively manipulated in a transport network. Through these waypoints, we can visually explore the OD patterns for the trajectories that successively pass through them. Second, we devise an efficient hashing-based query method to perform real-time waypoints-constrained trajectory filtering. Third, we develop the *waypoints-constrained OD view* to present the spatial-, temporal- and path-related information of the OD patterns. Lastly, we perform two case studies on the EZLink data, and conduct an interview with several transport researchers to examine our visual analytics method.

Chapter 6: Visual Analytics for PTS Mobility

This chapter visualizes and explores commuter mobility in a PTS. After we define the problem and the analytical tasks, we then introduce and construct a PTS mobility model that characterizes the commuter mobility, and derive methods to estimate various mobility-related factors, including waiting time, riding time, transfer time and travel efficiency, from the EZLink data. Our

visual analytic interface is an integrated solution with three visualization modules: *isochrone map view*, *isotime flow map view*, and *OD-pair journey view*, enabling us to efficiently perform the analytical tasks concerning time-efficient journeys originated from a given starting location. Particularly, the *isotime flow map view* is a novel visualization strategy, which linearizes a flow map in a parallel isotime layout, thereby presenting clear and smooth pathways from the given origin to destinations as well as maximizing the visualization and comparison of various mobility-related factors along the routes. To come up with this design, we also explore and compare two other design alternatives. In the end, we also explore two case studies with the transport researchers, and present their expert feedbacks on the interface design.

Besides, Chapter 2 presents an overview of state-of-the-art visualization and analysis techniques in the scope of this thesis. Chapter 3 presents a detailed description of the input dataset, including its characteristics and examples of its applications in transport domain.

7.2 Contributions

The thesis has demonstrated the usefulness of *information visualization reference model* [28] in developing effective visual analytics. The reference model suggests the framework of successfully developing information visualization: identifying the user tasks, extracting knowledge from data, mapping the knowledge to visual structures, and constructing the views from the visual structures. By closely following the model, we have successfully developed three visual analytics to address transportation researchers' tasks. Besides, the thesis also demonstrated the effectiveness of visual analytics in data exploration process: With appropriately designed visual analytics, domain experts can progressively refine and evaluate their analysis results.

The major contributions of this thesis lie in two perspectives.

First, from the perspective of visual analytics, we have made the following achievements: 1) related state-of-the-art information visualization techniques, particularly the visual analytics

for movement data and visualizations of urban traffic, have been summarized in Chapter 2; 2) a family of novel visual representations has been proposed to visualize the input transport data, i.e., the *interchange circos diagram* in Chapter 4, the *waypoints-constrained OD view* in Chapter 5 and an integrated visual analytics with *isochrone map view*, *isotime flow map view*, and *OD-pair journey view* in Chapter 6; 3) correspondingly, a family of user interactions has been developed to allow the transport researchers to interactively explore the visualizations; and 4) the applicability and limitations of these visual analytics have also been discussed.

Second, from the perspective of application domain, i.e., transport in this thesis, we have achieved the following successes: 1) a set of analytical tasks has been identified from transport researchers, i.e., to reveal commuter interchange patterns at junction nodes in a traffic network (Chapter 4), to explore the OD patterns associated with commuter trajectories successively passing through users-specified entry and exit waypoints in a transport network (Chapter 5), and to present mobility information of a PTS (Chapter 6); 2) with effective data modeling and mining techniques, the task-related information has been successfully extracted from the input public transport dataset; and 3) the extracted information has been successfully depicted with novel visualization methods in an efficient and intuitive way, such as the interchange circos diagram and isotime flow map view.

7.3 Future Work

There is still space to improve the current visual analytics systems. For instance, the mobility visual analytics system (Chapter 6) presents only waiting time, riding time, transfer time and travel efficiency along time-efficient journeys. In the future, we would like to explore other transport attributes such as vehicle capacity and commuter composition (e.g., senior, student, and disabled). Moreover, we also plan to extend the system for real-time analysis of PTS mobility, so that we may deliver adaptive journey planning for commuters.

The thesis has mainly explored the visual analytics for three different aspects of a PTS, i.e., the interchange patterns, waypoint-constrained OD patterns, and mobility. In other words, we have only addressed *some* of the analytical tasks that are interesting for transport researchers. There remain open researches for visualization communities to develop visual analytics for the researchers to analyze the urban public transport data. One example would be to develop a visual analytical system to explore and analyze commuter route choice model. In more detail, given an OD pair, there maybe multiple routes selected by commuters due to their different traveling preferences. We would like to extract all these possible routes from the input dataset, and explore and visualize the factors affecting the commuter choice.

Besides, visual fusion of the public transport data with other kinds of mega-city big data, e.g., social media data, may provide more insights for explaining commuters' behaviors. For instance, Krueger et al. [73] showed how semantic insights can be gained by enriching trajectory data with POI information extracted from Foursquare data. It is worth to explore the possibilities in integrating these different kinds of dataset.

What's more, we believe that nowadays the availability of wide range visual displays, e.g., from small screens on mobile phones to extra-large displays, calls for new visualization and interaction methods for exploring the public transport data. We would also like to work in this direction as a long-term work.

References

- [1] Nobbir Ahmed and Harvey J. Miller. Time–space transformations of geographic space for exploring, analyzing and visualizing transportation systems. *Journal of Transport Geography*, 15(1):2–17, January 2007.
- [2] Wolfgang Aigner, Silvia Miksch, Wolfgang Müller, Heidrun Schumann, and Christian Tominski. Visual methods for analyzing time-oriented data. *IEEE Transactions on Visualization and Computer Graphics*, 14(1):47–60, January 2008.
- [3] Wolfgang Aigner, Silvia Miksch, Heidrun Schumann, and Christian Tominski. *Visualization of time-oriented data*. Springer, 2011.
- [4] Bilal Alsallakh, Eduard Gröller, Silvia Miksch, and Martin Suntinger. Contingency Wheel: Visual Analysis of Large Contingency Tables. In *International Workshop on Visual Analytics (EuroVA’11)*, pages 53–56. Eurographics Association, 2011.
- [5] Fereshteh Amini, Sebastien Rufige, Zahid Hossain, Quentin Ventura, Pourang Irani, and Michael J. McGuffin. The Impact of Interactivity on Comprehending 2D and 3D Visualizations of Movement Data. *IEEE Transactions on Visualization and Computer Graphics*, 21(1):122 – 135, 2014.
- [6] Gennady Andrienko and Natalia Andrienko. Spatio-temporal aggregation for visual analysis of movements. In *IEEE Symposium on Visual Analytics Science and Technology, VAST’08*, pages 51–58, October 2008.
- [7] Gennady Andrienko, Natalia Andrienko, Peter Bak, Daniel Keim, and Stefan Wrobel. *Visual Analytics of Movement*. Springer, 2013.

- [8] Gennady Andrienko, Natalia Andrienko, Jason Dykes, Sara Irina Fabrikant, and Monica Wachowicz. Geovisualization of Dynamics, Movement and Change: Key Issues and Developing Approaches in Visualization Research. *Information Visualization*, 7:173–180, 2008.
- [9] Gennady Andrienko, Natalia Andrienko, Salvatore Rinzivillo, Mirco Nanni, Dino Pedreschi, and Fosca Giannotti. Interactive visual clustering of large collections of trajectories. In *IEEE Symposium on Visual Analytics Science and Technology, VAST’09*, pages 3–10, Atlantic City, NJ, October 2009.
- [10] Natalia Andrienko and Gennady Andrienko. Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics*, 17(2):205–219, February 2011.
- [11] Natalia Andrienko and Gennady Andrienko. Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, 12(1):3–24, September 2012.
- [12] Natalia Andrienko, Gennady Andrienko, Nikos Pelekis, and Stefano Spaccapietra. Basic concepts of movement data. In *Mobility, Data Mining and Privacy*, pages 15–38. Springer Berlin Heidelberg, 2008.
- [13] Luc Anselin. What is Special About Spatial Data? Alternative Perspectives on Spatial Data Analysis. In *Symposium on Spatial Statistics, Past, Present and Future*, pages 63–77. Department of Geography, Syracuse University, 1989.
- [14] Arthur Appel, F. James Rohlf, and Arthur J. Stein. The haloed line effect for hidden line elimination. *SIGGRAPH Computer Graphics*, 13(2):151–157, August 1979.
- [15] Land Transport Authority. Land Transport Master Plan 2013. Technical report, Singapore, 2013.
- [16] Kay W. Axhausen, Claudia Dolci, Ph Fröhlich, Milena Scherer, and Alessandro Carosio. Constructing Time-Scaled Maps: Switzerland from 1950 to 2000. *Transport Reviews*, 28(3):391–413, 2008.

REFERENCES

- [17] Ricardo Barrero, Joeri Van Mierlo, and Xavier Tackoen. Energy savings in public transport. *IEEE Vehicular Technology Magazine*, 3(3):26–36, 2008.
- [18] D. J. Bartram. Comprehending spatial information: The relative efficiency of different methods of presenting information about bus routes. *Journal of Applied Psychology*, 65(1):103–110, February 1980.
- [19] Richard A. Becker, Stephen G. Eick, and Allan R. Wilks. Visualizing network data. *IEEE Transactions on Visualization and Computer Graphics*, 1(1):16–28, March 1995.
- [20] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3: Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, December 2011.
- [21] Ilya Boyandin. *Visualization of Temporal Origin-Destination Data*. Doctoral Thesis, University of Fribourg, Switzerland, 2013.
- [22] Ilya Boyandin, Enrico Bertini, Peter Bak, and Denis Lalanne. Flowstrates: An Approach for Visual Exploration of Temporal Origin-Destination Data. *Computer Graphics Forum*, 30(3):971–980, June 2011.
- [23] Ilya Boyandin, Enrico Bertini, and Denis Lalanne. A Qualitative Study on the Exploration of Temporal Changes in Flow Maps with Animation and Small-Multiples. *Computer Graphics Forum*, 31(3pt2):1005–1014, June 2012.
- [24] Runar Brännlund, Tarek Ghalwash, and Jonas Nordström. Increased energy efficiency and the rebound effect: Effects on consumption and emissions. *Energy Economics*, 29(1):1–17, January 2007.
- [25] Dirk Brockmann, Lars Hufnagel, and Theo Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, January 2006.
- [26] Michael Burch, Andreas Kull, and Daniel Weiskopf. AOI rivers for visualizing dynamic eye gaze frequencies. *Computer Graphics Forum*, 32(3pt3):281–290, 2013.
- [27] Francesco Calabrese, Mi Diao, Giusy Di Lorenzo, Joseph Ferreira Jr., and Carlo Ratti. Understanding individual mobility patterns from urban sensing data: A mobile phone

- trace example. *Transportation Research Part C: Emerging Technologies*, 26:301–313, 2013.
- [28] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.
- [29] M. Sheelagh T. Carpendale, David J. Cowperthwaite, and F. David Fracchia. Distortion Viewing Techniques for 3-Dimensional Data. In *IEEE Symposium on Information Visualization*, InfoVis’96, pages 46–53, San Francisco, CA, October 1996.
- [30] Qianwen Chao, Jingjing Shen, and Xiaogang Jin. Video-based personalized traffic learning. *Graphical Models*, 75(6):305–317, November 2013.
- [31] Ding Chu, David A Sheets, Ye Zhao, Yingyu Wu, Jing Yang, Maogong Zheng, and George Chen. Visualizing Hidden Themes of Taxi Movement with Semantic Transformation. In *IEEE Pacific Visualization Symposium*, PacificVis’14, pages 137–144, Yokohama, 2014.
- [32] Pio Claudio and Sung-Eui Yoon. Metro Transit-Centric Visualization for City Tour Planning. *Computer Graphics Forum*, 33(3):271–280, June 2014.
- [33] Vehlow Corinna, Fabian Beck, and Daniel Weiskopf. The State of the Art in Visualizing Group Structures in Graphs. In *Eurographics Conference on Visualization (EuroVis) - STARs*, pages 21–40, Cagliari, Italy, 2015. The Eurographics Association.
- [34] Weiwei Cui, Hong Zhou, Huamin Qu, Pak Chung Wong, and Xiaoming Li. Geometry-based edge clustering for graph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1277–1284, 2008.
- [35] Somayeh Dodge, Robert Weibel, and Ehsan Forootan. Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems*, 33(6):419–434, 2009.
- [36] Somayeh Dodge, Robert Weibel, and Anna-Katharina Lautenschütz. Towards a taxonomy of movement patterns. *Information Visualization*, 7(3):240–252, 2008.
- [37] Harish Doraiswamy, Nivan Ferreira, Theodoros Damoulas, Juliana Freire, and Cláudio T. Silva. Using Topological Analysis to Support Event-Guided Exploration in

REFERENCES

- Urban Data. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2634–2643, 2014.
- [38] Alexander Erath, Pieter J. Fourie, Michael van Eggermond, Sergio A. Ordóñez, Artem Chakirov, and Kay W. Axhausen. Large-scale, agent-based transport demand model for Singapore. In *International Conference on Travel Behaviour Research (IATBR)*, pages 1–39, Toronto, June 2012. International Association for Travel Behaviour Research.
- [39] Ozan Ersoy, Christophe Hurter, Fernando Vieira Paulovich, Gabriel Cantareiro, and Alexandru Telea. Skeleton-based edge bundling for graph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2364–2373, December 2011.
- [40] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3):37–54, 1996.
- [41] Nivan Ferreira, James T. Klosowski, Carlos E. Scheidegger, and Cláudio T. Silva. Vector Field k-Means: Clustering Trajectories by Fitting Multiple Vector Fields. *Computer Graphics Forum*, 32(3pt2):201–210, June 2013.
- [42] Nivan Ferreira, Jorge Poco, Huy T. Vo, Juliana Freire, and Cláudio T. Silva. Visual exploration of big spatio-temporal urban data: A study of New York city taxi trips. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2149–2158, December 2013.
- [43] Mark Gahegan. Four barriers to the development of effective exploratory visualisation tools for the geosciences. *International Journal of Geographical Information Science*, 13(4):289–309, 1999.
- [44] Ken Garland. *Mr Beck’s underground map*. Capital Transport Publishing, 1994.
- [45] C. E. Gehlke and Katherine Biehl. Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association*, 29(185):168–170, March 1934.
- [46] Mohammad Ghoniem, J Fekete, and Philippe Castagliola. A comparison of the readability of graphs using node-link and matrix-based representations. In *IEEE Symposium on Information Visualization*, InfoVis’04, pages 17–24, 2004.

REFERENCES

- [47] Fosca Giannotti, Mirco Nanni, Dino Pedreschi, and Fabio Pinelli. Trajectory pattern mining. In *International Conference on Knowledge Discovery and Data Mining*, KDD'07, pages 330–339, San Jose, California, August 2007.
- [48] Tovi Grossman, Daniel Wigdor, and Ravin Balakrishnan. Exploring and Reducing the Effects of Orientation on Text Readability in Volumetric Displays. In *SIGCHI Conference on Human Factors in Computing Systems*, CHI'07, pages 483–492, San Jose, CA, May 2007.
- [49] Joachim Gudmundsson, Marc van Kreveld, and Bettina Speckmann. Efficient Detection of Patterns in 2D Trajectories of Moving Points. *Geoinformatica*, 11(2):195–215, June 2007.
- [50] Diansheng Guo. Flow mapping and multivariate visualization of large spatial interaction data. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1041–1048, December 2009.
- [51] Diansheng Guo, Jin Chen, Alan M. MacEachren, and Ke Liao. A visualization system for space-time and multivariate patterns (VIS-STAMP). *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1461–1474, December 2006.
- [52] Diansheng Guo and Xi Zhu. Origin-Destination Flow Data Smoothing and Mapping. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2043–2052, 2014.
- [53] Diansheng Guo, Xi Zhu, Hai Jin, Peng Gao, and Clio Andris. Discovering Spatial Patterns in Origin-Destination Mobility Data. *Transactions in GIS*, 16(3):411–429, 2012.
- [54] Hanqi Guo, Zuchao Wang, Bowen Yu, Huijing Zhao, and Xiaoru Yuan. TripVista: Triple perspective visual trajectory analytics and its application on microscopic traffic data at a road intersection. In *IEEE Pacific Visualization Symposium*, PacificVis'11, pages 163–170, Hong Kong, March 2011. IEEE.
- [55] Torsten Hägerstrand. What About People in Regional Science? *Papers of the Regional Science Association*, 24(1):6–21, 1970.
- [56] Susan Havre, Beth Hetzler, and Lucy Nowell. ThemeRiver: Visualizing theme changes over time. In *IEEE Symposium on Information Visualization*, InfoVis'00, pages 115–123, Salt Lake City, UT, 2000.

- [57] Jeffrey Heer, Nicholas Kong, and Maneesh Agrawala. Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations. In *SIGCHI Conference on Human Factors in Computing Systems*, CHI'09, pages 1303–1312, Boston, MA, 2009. ACM.
- [58] Dirk Helbing. Traffic and related self-driven many-particle systems. *Reviews of Modern Physics*, 73(4):1067–1141, December 2001.
- [59] Danny Holten and Jarke J. van Wijk. Visual Comparison of Hierarchically Organized Data. *Comput. Graph. Forum*, 27(3):759–766, 2008.
- [60] Danny Holten and Jarke J. van Wijk. Force-Directed Edge Bundling for Graph Visualization. *Computer Graphics Forum*, 28(3):983–990, 2009.
- [61] Kathleen Hornsby and Max J. Egenhofer. Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence*, 36(1–2):177–194, 2002.
- [62] Christophe Hurter, Ozan Ersoy, and Alexandru Telea. Graph bundling by kernel density estimation. *Computer Graphics Forum*, 31(3pt1):865–874, June 2012.
- [63] Christophe Hurter, Benjamin Tissoires, and Stéphane Conversy. Accumulation as a tool for efficient visualization of geographical and temporal data. In *AGILE Workshop Geospatial Visual Analytics: Focus on Time*, 2009.
- [64] Christophe Hurter, Benjamin Tissoires, and Stéphane Conversy. FromDaDy: Spreading aircraft trajectories across views to support iterative queries. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1017–1024, November 2009.
- [65] David Husch and John Albeck. *Intersection Capacity Utilization: Evaluation Procedures for Intersections and Interchanges*. Trafficware, January 2003.
- [66] Victoria Interrante and Chester Grosch. Visualizing 3D Flow. *IEEE Computer Graphics and Applications*, 18(4):49–53, August 1998.
- [67] Jerald Jariyasunant, Eric Mai, and Raja Sengupta. Algorithm for Finding Optimal Paths in a Public Transit Network with Real-Time Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2256(1):34–42, 2011.

- [68] Shan Jiang, Joseph Ferreira, and Marta C. Gonzalez. Discovering urban spatial-temporal structure from human activity patterns. In *ACM SIGKDD International Workshop on Urban Computing*, UrbComp'12, pages 95–102, 2012.
- [69] Thomas Kapler and William Wright. GeoTime Information Visualization. In *IEEE Symposium on Information Visualization*, InfoVis'04, pages 25–32, Austin, TX, October 2004.
- [70] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. Visual Analytics: Definition, Process, and Challenges. In *Information Visualization: Human-Centered Issues and Perspectives*, volume 4950 of *Lecture Notes in Computer Science*, pages 154–175. Springer Berlin Heidelberg, 2008.
- [71] Daniel Keim, Jörn Kohlhammer, Geoffrey Ellis, and Florian Mansmann. *Mastering the Information Age: Solving Problems with Visual Analytics*. The Eurographics Association, 2010.
- [72] Per Ola Kristensson, Nils Dahlbäck, Daniel Amundi, Marius Björnstad, Hanna Gillberg, Jonas Haraldsson, Ingrid Mårtensson, and Josefine Sthål. An Evaluation of Space Time Cube Representation of Spatiotemporal Patterns. *IEEE Transactions on Visualization and Computer Graphics*, 15(4):696–702, August 2009.
- [73] Robert Krüger, Dennis Thom, and Thomas Ertl. Semantic Enrichment of Movement Behavior with Foursquare - A Visual Analytics Approach. *IEEE Transactions on Visualization and Computer Graphics*, 21(8):903 – 915, 2015.
- [74] Martin Krzywinski, Jacqueline Schein, Inanc Birol, Joseph Connors, Randy Gascoyne, Doug Horsman, Steven J. Jones, and Marco A. Marra. Circos: an information aesthetic for comparative genomics. *Genome Research*, 19(9):1639–1645, 2009.
- [75] Mei-Po Kwan. Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research Part C: Emerging Technologies*, 8(1):185–203, February 2000.
- [76] Antoine Lambert, Romain Bourqui, and David Auber. Winding Roads: Routing edges into bundles. *Computer Graphics Forum*, 29(3):853–862, 2010.

REFERENCES

- [77] Patrick Laube and Stephan Imfeld. Analyzing relative motion within groups of trackable moving point objects. In *International Conference on Geographic Information Science, GIScience'02*, pages 132–144, London, 2002.
- [78] Patrick Laube, Stephan Imfeld, and Robert Weibel. Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science*, 19(6):639–668, 2005.
- [79] Patrick Laube and Ross S. Purves. An approach to evaluating motion pattern detection techniques in spatio-temporal data. *Computers, Environment and Urban Systems*, 30(3):347–374, May 2006.
- [80] Patrick Laube, Marc van Kreveld, and Stephan Imfeld. Finding REMO - Detecting relative motion patterns in geospatial lifelines. In *Developments in Spatial Data Handling: 11th International Symposium on Spatial Data Handling*, pages 201–214, 2004.
- [81] Richard Layard and Stephen Glaister. *Cost-Benefit Analysis*. Cambridge University Press, June 1994.
- [82] Todd Litman. *Evaluating accessibility for transportation planning*. Victoria Transport Policy Institute, 2008.
- [83] He Liu, Yuan Gao, Lu Lu, Siyuan Liu, Huamin Qu, and Lionel M. Ni. Visual Analysis of Route Diversity. In *IEEE Conference on Visual Analytics Science and Technology, VAST'11*, pages 171–180, Providence, RI, October 2011.
- [84] Liang Liu, Assaf Biderman, and Carlo Ratti. Urban Mobility Landscape: Real Time Monitoring of Urban Mobility Patterns. In *International Conference on Computers in Urban Planning and Urban Management, CUPUM'09*, pages 1–16, 2009.
- [85] Siyuan Liu, Jiansu Pu, Qiong Luo, Huamin Qu, Lionel M. Ni, and Ramayya Krishnan. VAIT: A Visual Analytics System for Metropolitan Transportation. *IEEE Transactions on Intelligent Transportation Systems*, 14(4):1586–1596, May 2013.
- [86] Suxia Liu and Xuan Zhu. Accessibility analyst: an integrated GIS tool for accessibility analysis in urban transportation planning. *Environment and Planning B: Planning and Design*, 31(1):105–124, 2004.

- [87] Alan M. MacEachren and Menno-Jan Kraak. Research challenges in geovisualization. *Cartography and Geographic Information Science*, 28(1):3–12, 2001.
- [88] Tobias Meilinger, Christoph Hölscher, Simon J Büchner, and Martin Brösamle. How much information do you need? Schematic maps in wayfinding and self localisation. In *Spatial Cognition V - Reasoning, Action, Interaction*, pages 381–400. Springer, 2007.
- [89] Mark Monmonier. *How to lie with maps*. University of Chicago Press, second edition edition, May 1996.
- [90] Till Nagel, Erik Duval, Andrew Vande Moere, Kristian Kloeckl, and Carlo Ratti. Sankey arcs – Visualizing edge weights in path graphs. In *Eurographics Conference on Visualization*, EuroVis’12, pages 55–59, Vienna, June 2012.
- [91] Martin Nöllenburg and Alexander Wolff. Drawing and labeling high-quality metro maps by mixed-integer programming. *IEEE Transactions on Visualization and Computer Graphics*, 17(5):626–641, 2011.
- [92] Stan Openshaw. *The Modifiable Areal Unit Problem*. Geo Books, Norwick, UK, 1983.
- [93] Juan de Dios Ortuzar and Luis G. Willumsen. *Modelling Transport*. Wiley, March 2011.
- [94] Mark Ovenden. *Metro maps of the world*. Capital Transport, 2005.
- [95] Nikos Pelekis, Ioannis Kopanakis, Gerasimos Marketos, Irene Ntoutsis, Gennady Andrienko, and Yannis Theodoridis. Similarity search in trajectory databases. In *International Symposium on Temporal Representation and Reasoning*, pages 129–140, Alicante, June 2007.
- [96] Donna J. Peuquet. It’s about time: a conceptual framework for the representation of temporal dynamics in geographic information systems. *Annals of the Association of American Geographers*, 84(3):441–461, 1994.
- [97] Donna J. Peuquet. *Representations of space and time*. Guilford Press, 2002.
- [98] Doantam Phan, Ling Xiao, Ron Yeh, Pat Hanrahan, and Terry Winograd. Flow map layout. In *IEEE Symposium on Information Visualization*, InfoVis’05, pages 219–224, 2005.

REFERENCES

- [99] Evangelia Pyrga, Frank Schulz, Dorothea Wagner, and Christos Zaroliagis. Efficient models for timetable information in public transportation systems. *Journal of Experimental Algorithmics*, 12:1–39, 2008.
- [100] Craig W. Reynolds. Flocks, herds and schools: A distributed behavioral model. *SIGGRAPH Computer Graphics*, 21(4):25–34, 1987.
- [101] Patrick Riehmann, Manfred Hanfler, and Bernd Froehlich. Interactive sankey diagrams. In *IEEE Symposium on Information Visualization*, InfoVis’05, pages 233–240, October 2005.
- [102] Salvatore Rinzivillo, Dino Pedreschi, Mirco Nanni, Fosca Giannotti, Natalia Andrienko, and Gennady Andrienko. Visually driven analysis of movement data by progressive clustering. *Information Visualization*, 7(3):225–239, 2008.
- [103] George Robertson, Roland Fernandez, Danyel Fisher, Bongshin Lee, and John Stasko. Effectiveness of animation in trend visualization. *IEEE Trans. Vis. Comp. Graphics*, 14(6):1325 – 1332, 2008.
- [104] Jean-Paul Rodrigue, Claude Comtois, and Brian Slack. *The Geography of Transport Systems*. Routledge, New York, 3rd edition, 2013.
- [105] Roeland Scheepens, Niels Willems, Huub van de Wetering, Gennady Andrienko, Natalia Andrienko, and Jarke J. van Wijk. Composite Density Maps for Multivariate Trajectories. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2518–2527, December 2011.
- [106] Roeland Scheepens, Niels Willems, Huub van de Wetering, and Jarke J. van Wijk. Interactive visualization of multivariate trajectory data with density maps. In *IEEE Pacific Visualization Symposium*, PacificVis’11, pages 147–154, Hong Kong, March 2011.
- [107] Roeland Scheepens, Niels Willems, Huub van de Wetering, and Jarke J. van Wijk. Interactive density maps for moving objects. *IEEE Computer Graphics and Applications*, 32(1):56–66, February 2012.
- [108] Jason Sewall, Jur van den Berg, Ming C. Lin, and Dinesh Manocha. Virtualized Traffic: Reconstructing Traffic Flows from Discrete Spatio-Temporal Data. *IEEE Transactions on Visualization and Computer Graphics*, 17(1):26–37, January 2011.

REFERENCES

- [109] Jason Sewall, David Wilkie, and Ming C. Lin. Interactive hybrid simulation of large-scale traffic. *ACM Transactions on Graphics*, 30(6):135:1–135:12, December 2011.
- [110] Jason Sewall, David Wilkie, Paul Merrell, and Ming C. Lin. Continuum Traffic Simulation. *Computer Graphics Forum*, 29(2):439–448, May 2010.
- [111] Jingjing Shen and Xiaogang Jin. Detailed Traffic Animation for Urban Road Networks. *Graphical Models*, 74(5):265–282, September 2012.
- [112] Eihan Shimizu. Time-space mapping based on topological transformation of physical map. In *World Conference on Transport Research*, volume 1, pages 219–230, Lyons, 1993.
- [113] Eihan Shimizu and Ryo Inoue. A new algorithm for distance cartogram construction. *International Journal of Geographical Information Science*, 23(11):1453–1470, 2009.
- [114] Bernard W. Silverman. Density Estimation for Statistics and Data Analysis. In *Mono-graphs on Statistics and Applied Probability*. Chapman and Hall, London, 1986.
- [115] Klaus Spiekermann and Michael Wegener. The shrinking continent: new time-space maps of Europe. *Environment and Planning B: Planning and Design*, 21(6):653–673, 1994.
- [116] Guodao Sun, Yingcai Wu, Ronghua Liang, and Shixia Liu. A Survey of Visual Analytics Techniques and Applications: State-of-the-Art Research and Future Challenges. *Journal of Computer Science and Technology*, 28(5):852–867, September 2013.
- [117] O. Z. Tamin and L. G. Willumsen. Transport demand model estimation from traffic counts. *Transportation*, 16(1):3–26, March 1989.
- [118] James J. Thomas and Kristin A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center, 2004.
- [119] Waldo R. Tobler. A computer movie simulating urban growth. *Economic Geography*, 46:234–240, June 1970.
- [120] Waldo R. Tobler. A model of geographical movement. *Geographical Analysis*, 13(1):1–20, 1981.

REFERENCES

- [121] Waldo R. Tobler. Experiments in migration mapping by computer. *The American Cartographer*, 14(2):155–163, 1987.
- [122] Christian Tominski, Petra Schulze-Wollgast, and Heidrun Schumann. 3D Information Visualization for Time Dependent Data on Maps. In *IEEE Symposium on Information Visualization*, InfoVis’05, pages 175–181, Minneapolis, MN, July 2005.
- [123] Christian Tominski, Heidrun Schumann, Gennady Andrienko, and Natalia Andrienko. Stacking-Based Visualization of Trajectory Attribute Data. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2565 – 2574, December 2012.
- [124] Edward R. Tufte. *Envisioning Information*. Graphics Press, 1990.
- [125] Barbara Tversky, Julie Bauer Morrison, and Mireille Betrancourt. Animation: can it facilitate? *International Journal of Human-Computer Studies*, 57(4):247–262, 2002.
- [126] Corinna Vehlow, Fabian Beck, Patrick Auwärter, and Daniel Weiskopf. Visualizing the evolution of communities in dynamic graphs. *Comput. Graph. Forum*, 34(1):277–288, 2015.
- [127] Kevin Verbeek, Kevin Buchin, and Bettina Speckmann. Flow map layout via spiral trees. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2536–2544, December 2011.
- [128] Jose M. Viegas. Making urban road pricing acceptable and effective: searching for quality and equity in urban mobility. *Transport Policy*, 8(4):289–294, 2001.
- [129] Alan M. Voorhees. A general theory of traffic movement. *Transportation*, 40(6):1105–1116, 2013.
- [130] Katerina Vrotsou, Halldor Janetzko, Carlo Navarra, Georg Fuchs, David Spretke, Florian Mansmann, Natalia Andrienko, and Gennady Andrienko. SimpliFly: A Methodology for Simplification and Thematic Enhancement of Trajectories. *IEEE Transactions on Visualization and Computer Graphics*, 21(1):107 – 121, July 2014.
- [131] Fei Wang, Wei Chen, Feiran Wu, Ye Zhao, Han Hong, Tianyu Gu, Long Wang, Ronghua Liang, and Hujun Bao. A Visual Reasoning Approach for Data-driven Transport Assess-

REFERENCES

- ment on Urban Road. In *IEEE Conference on Visual Analytics Science and Technology, VAST'14*, pages 103 – 112, Paris, France, 2014.
- [132] Pu Wang, Timothy Hunter, Alexandre M. Bayen, Katja Schechtner, and Marta C. González. Understanding Road Usage Patterns in Urban Areas. *Scientific Reports*, 2(1001), 2012. Article 1001.
- [133] Yu-Shuen Wang and Ming-Te Chi. Focus+context metro maps. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2528–2535, December 2011.
- [134] Zuchao Wang, Min Lu, Xiaoru Yuan, Junping Zhang, and Huub van de Wetering. Visual traffic jam analysis based on trajectory data. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2159–2168, December 2013.
- [135] Zuchao Wang, Tangzhi Ye, Min Lu, Xiaoru Yuan, Huamin Qu, Jacky Yuan, and Qianliang Wu. Visual Exploration of Sparse Traffic Trajectory Data. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1813–1822, December 2014.
- [136] Marc Weber, Marc Alexa, and Wolfgang Müller. Visualizing time-series on spirals. In *IEEE Symposium on Information Visualization, InfoVis'01*, pages 7–13, 2001.
- [137] Wikipedia. Isochrone map, 2014.
- [138] David Wilkie, Jason Sewall, and Ming C. Lin. Transforming GIS Data into Functional Road Models for Large-Scale Traffic Simulation. *IEEE Transactions on Visualization and Computer Graphics*, 18(6):890–901, June 2012.
- [139] David Wilkie, Jason Sewall, and Ming C. Lin. Flow Reconstruction for Data-Driven Traffic Animation. *ACM Transactions on Graphics*, 23(4):89:1–89:10, July 2013.
- [140] Niels Willems, Roeland Scheepens, Huub van de Wetering, and Jarke J. van Wijk. Visualization of Vessel Traffic. In *Situation Awareness with Systems of Systems*, pages 73–87. Springer New York, January 2013.
- [141] Niels Willems, Huub van de Wetering, and Jarke J. van Wijk. Visualization of vessel movements. *Computer Graphics Forum*, 28(3):959–966, 2009.
- [142] Jo Wood, Jason Dykes, and Aidan Slingsby. Visualization of origins, destinations and flows with OD maps. *The Cartographic Journal*, 47(2):117–129, May 2010.

REFERENCES

- [143] Jo Wood, Aidan Slingsby, and Jason Dykes. Visualizing the Dynamics of London’s Bicycle-Hire Scheme. *Cartographica*, 46(4):239–251, 2011.
- [144] Hsiang-Yun Wu, Shigeo Takahashi, Daichi Hirono, Masatoshi Arikawa, Chun-Cheng Lin, and Hsu-Chun Yen. Spatially Efficient Design of Annotated Metro Maps. *Computer Graphics Forum*, 32(3pt3):261–270, 2013.
- [145] Panpan Xu, Yingcai Wu, Enxun Wei, Tai-Quan Peng, Shixia Liu, Jonathan J. H. Zhu, and Huamin Qu. Visual analysis of topic competition on social media. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2012 – 2021, December 2013.
- [146] Jing Yuan, Yu Zheng, and Xing Xie. Discovering regions of different functions in a city using human mobility and POIs. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD’12, pages 186–194, 2012.
- [147] Jing Yuan, Yu Zheng, Xing Xie, and Guangzhong Sun. T-Drive: Enhancing Driving Directions with Taxi Drivers’ Intelligence. *IEEE Transactions on Knowledge and Data Engineering*, 25(1):220–232, January 2013.
- [148] Jing Yuan, Yu Zheng, Chengyang Zhang, Wenlei Xie, Xing Xie, Guangzhong Sun, and Yan Huang. T-drive: Driving directions based on taxi trajectories. In *International Conference on Advances in Geographic Information Systems*, GIS’10, pages 99–108, 2010.
- [149] Jing Yuan, Yu Zheng, Liuhang Zhang, Xing Xie, and Guangzhong Sun. Where to find my next passenger. In *International Conference on Ubiquitous Computing*, Ubi-Comp’11, pages 109–118, 2011.
- [150] Wei Zeng, Chi-Wing Fu, Stefan Müller Arisona, Alexander Erath, and Huamin Qu. Visualizing Mobility of Public Transportation System. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1833–1842, December 2014.
- [151] Wei Zeng, Chi-Wing Fu, Stefan Müller Arisona, and Huamin Qu. Visualizing Interchange Patterns in Massive Movement Data. *Computer Graphics Forum*, 32(3pt3):271–280, June 2013.

REFERENCES

- [152] Wei Zeng, Chi-Wing Fu, Stefan Müller Arisona, and Huamin Qu. Visualizing waypoints-constrained origin-destination pattern for massive urban transportation data. *Computer Graphics Forum*, 2015. minor revision.
- [153] Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. Urban Computing: Concepts, Methodologies, and Applications. *ACM Transactions on Intelligent Systems and Technology*, 5(3):38:1–38:55, October 2014.
- [154] Xi Zhu and Diansheng Guo. Mapping Large Spatial Flow Data with Hierarchical Clustering. *Transactions in GIS*, 18(3):421–435, May 2014.
- [155] Xuan Zhu and Suxia Liu. Analysis of the impact of the MRT system on accessibility in Singapore using an integrated GIS tool. *Journal of Transport Geography*, 12(2):89–101, 2004.

Publications

- **Wei Zeng**, Chi-Wing Fu, Stefan Müller Arisona, Alexander Erath and Huamin Qu, “Visualizing Waypoints-Constrained Origin-Destination Pattern for Massive Urban Transportation Data”, *Computer Graphics Forum*, in print.
- **Wei Zeng**, Chi-Wing Fu, Stefan Müller Arisona, Alexander Erath and Huamin Qu, “Visualizing Mobility of Public Transportation System”, *Proceedings of IEEE Visual Analytics Science and Technology (VAST, IEEE VIS 2014)*, selected to be published in *Transactions on Visualization and Computer Graphics (TVCG)*, vol. 20, num. 12, pp. 1833-1842, 2014.
- Afian Anwar, **Wei Zeng**, Stefan Müller Arisona, “The Time Space Diagram Revisited”, *Proceedings of Transportation Research Board 93th Annual Meeting (TRB’14)*, selected to be published in *Transportation Research Record: Journal of the Transportation Research Board*, no. 2442, 2014. (top conference in transportation since 1920)
- **Wei Zeng**, Xianfeng Huang, Stefan Müller Arisona and Ian Vince McLoughlin, “Classifying Watermelon Ripeness by Analysing Acoustic Signals Using Mobile Devices”, *Personal and Ubiquitous Computing*, vol. 18, issue 7, pp. 1753-1762, 2014.
- **Wei Zeng**, Chi-Wing Fu, Stefan Müller Arisona and Huamin Qu, “Visualizing Interchange Patterns in Massive Movement Data”, *Proceedings of the Eurographics Conference on Visualization (EuroVis’13)*, published in *Computer Graphics Forum (CGF)*, vol. 32, issue 3pt3, pp. 271-280, 2013.
- **Wei Zeng**, Chen Zhong, Afian Anwar, Stefan Müller Arisona and Ian Vince McLoughlin, “MetroBuzz: Interactive 3D Visualization of Spatiotemporal Data”, *International Conference on Computer & Information Science (ICCIS’12)*, vol. 1, pp. 143-147, 2012.

REFERENCES

- Chen Zhong, Tao Wang, **Wei Zeng**, and Stefan Müller Arisona, “Spatiotemporal Visualization: A Survey and Outlook”, *Digital Urban Modeling and Simulation*, vol. 242, pp. 299-317, 2012.