

ResearchOnline@JCU

This file is part of the following reference:

Wang, Ye (2017) *Visual analytics of dynamic higher order information*. PhD thesis, James Cook University.

Access to this file is available from:

<https://researchonline.jcu.edu.au/51660/>

The author has certified to JCU that they have made a reasonable effort to gain permission and acknowledge the owner of any third party copyright material included in this document. If you believe that this is not the case, please contact

*ResearchOnline@jcu.edu.au and quote
<https://researchonline.jcu.edu.au/51660/>*

Visual Analytics of Dynamic Higher Order Information



Ye Wang

Information Technology Academy
College of Business, Law and Governance
James Cook University
Cairns, QLD 4870, Australia

This dissertation is submitted for the degree of
Doctor of Philosophy

November 2017

Statement of Access

I, the undersigned, author of this work, understand that James Cook University will make this thesis available for use within the University Library and, via the Australian Digital Theses network, for use elsewhere. I understand that, as an unpublished work, a thesis has significant protection under the copyright act and; I do not wish to place any further restriction on access to this work.

Ye Wang

November 2017

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Ye Wang

November 2017

Acknowledgements

Due to the support and encouragement of many university staff, friends and family, and I am grateful to all those who have helped me complete my research.

I would like to thank my current supervisors Professor Ickjai Lee, Dr Kyungmi (Joanne) Lee and Dr Carrie Lui, for their support and assistance throughout this research.

I would also like to thank the staff and students of the Discipline of IT for providing an enjoyable and educational research environment.

Finally, I would like to thank my parents and family in China for supporting me throughout my research.

Publications

International Journals with full paper refereed

- J1:** Ye Wang and Ickjai Lee. “Visual Analytics of Higher Order Information for Trajectory Datasets”. *World Academy of Science, Engineering and Technology, International Journal of Environmental, Chemical, Ecological, Geological and Geophysical Engineering*, vol. 7, no. 12, pp: 966-971, 2014.
- J2:** Ye Wang, Kyungmi Lee, and Ickjai Lee. “Visual Analytical Tool for Multivariate Higher Order information for Emergency Management”. *Engineering Applications of Artificial Intelligence*, Elsevier, Submitted.
- J3:** Ye Wang, Kyungmi Lee, and Ickjai Lee. “DNA-chart Visual Tool of Topological Higher Order Information for Spatio-temporal Trajectory Datasets”. *Expert Systems with Applications*, Elsevier, Submitted.

International Conferences and Workshops with full paper refereed

- C1:** Ye Wang, Kyungmi Lee, and Ickjai Lee. “Visual Analytical Tool for Higher Order k -Means Clustering for Trajectory Data Mining”. In: Kang B., Bai Q. (eds) *AI2016: Advances in Artificial Intelligence - 29th Australasian Joint Conference*, Lecture Notes in Computer Science, vol. 9992, pp. 507-518, Springer, Hobart Australia, 2016.
- C2:** Ye Wang, Kyungmi Lee, and Ickjai Lee. “Multivariate Higher Order Information for Emergency Management Based on Tourism Trajectory Datasets”. In: Fujita H., Ali M., Selamat A., Sasaki J., Kurematsu M. (eds) *Trends in Applied Knowledge-Based Systems and Data Science - 29th International Conference on Industrial Engineering*

and Other Applications of Applied Intelligent Systems, Lecture Notes in Computer Science, vol.9799, pp. 725-736. Springer International Publishing, Morioka Japan, 2016.

- C3:** Ye Wang, Kyungmi Lee, and Ickjai Lee. “Treemap Representations of Higher Order Information for Moving Objects”. *10th International Conference on Intelligent Systems and Knowledge Engineering*, IEEE Computer Society, pp. 287-294, Taipei Taiwan, 2015.
- C4:** Ye Wang, Kyungmi Lee, and Ickjai Lee. “Directional higher order information for spatio-temporal trajectory dataset”. *Workshop on Spatial and Spatio-temporal Data Mining in conjunction with IEEE International Conference on Data Mining, ICDM Workshops*, IEEE Computer Society, pp. 35-42, Shenzhen China, 2014.
- C5:** Ye Wang, Kyungmi Lee, and Ickjai Lee. “Visual analytics of topological higher order information for emergency management based on tourism trajectory datasets”. *International Conference on Computational Science*, Procedia Computer Science, vol. 29, pp. 683-691, Cairns Australia, 2014.

Abstract

Nearest neighbour query and nearest region query providing higher-order information are the two most popular queries with spatio-temporal data. These types of higher-order information serve as the basis for “what-if” and “what-happen” scenarios. As our world turns into a more complex and uncertain environment with different emerging technologies, one of the challenging issues many decision-makers face is a data-rich, information-rich, but knowledge-poor situation. Multiple and dynamic datasets have been collected by various devices (sources). However, the time variant nature of these dynamic datasets will require different analytic tools and techniques to pre-process and analyse. Understanding and gaining insights into dynamic spatio-temporal data is of great importance in current ubiquitous environments, and visual analytics is a solid candidate to amplify user’s ability to detect expected patterns and to discover unexpected patterns from this type of dynamic spatio-temporal data.

Previous “what-if” analysis research mainly focused on single (univariate) datasets only. Existing tools that work with univariate spatio-temporal data do not have the data structure and cannot effectively support trajectory datasets. Moreover, related visual analytics methods to deal with dynamic spatio-temporal data, especially for higher-order information, are very limited. These methods are usually non-interactive that cannot support data exploration for users. The visualisation techniques used in existing methods also mainly focused on quantitative information but not qualitative information. In terms of different types of information, previous studies mostly focused on geometrical information with very limited attention in topological and directional information. There is also a lack of measurements and data mining techniques for decision makers to compare similarity/dissimilarity of higher-order information of trajectories.

The overall aim of this thesis is to investigate the application of visual analytics tools to provide various higher-order information with multivariate datasets of moving objects to provide decision support for users. First, we will design a robust data structure that can effectively support multivariate Point of Interests targets and multivariate dynamic moving objects of higher-order information. Second, the data structure needs to support not only the most widely used qualitative/quantitative geometrical higher-order information, but topological and directional information. Third, we will design and develop various visual analytical tools that help users to gain insights into the higher-order nature of trajectory data. Fourth, the proposed tools will be evaluated against effectiveness and efficiency. Fifth, we will define several dissimilarity measures for geometrical/topological/directional higher-order information and related data mining algorithms that return top- k similar trajectories. Lastly, we will present case studies with real data to validate the developed data structure and visual analytics tools.

The outcomes achieved by this research provide a number of contributions to the greater body of knowledge in data analytics and spatial temporal data analysis. We propose a set of visual analytics tools for exploring and analysing multivariate Points of Interest targets and multivariate dynamic moving objects of higher-order information. The contributions of this research include a new vector-based unified data structure, which allows datasets between Points of Interest and trajectories to be modelled and quantified. Two ways of data preprocessing of trajectory can be addressed when data is too dense or too sparse. In addition, there are five different types of visual analytics tools and three types of new measurement metrics for describing and comparing the similarity of higher-order information. These include the visual analytical tools, namely, higher-order parallel coordinates, higher-order spider, higher-order radar, higher-order treemaps, and higher-order DNA. For the measurement metrics, this research introduces three metrics, namely directional dissimilarity measurement, parallel coordinate multivariate measurement, and higher-order DNA impact factor. These visual tools and measurement metrics address the limitation of existing analytical tools to provide interactive tools for users to explore both qualitative and quantitative multivariate higher-order information with integration of all three different types of higher-order informa-

tion. A real-world emergency management case study will demonstrate the usefulness and applicability of these new visual analytics tools.

Abbreviations

DBR: Directional Balance Ratio

DRs: Directional Ratios

DDM: Directional Dissimilarity Measure

DSC: Directional Spider Chart

GIS Geographical Information Systems

GPS: Global Positioning System

HOCC: Higher Order CircumCircle

HODNA: Higher Order DNA

HODNAC: Higher Order DNA Chart

HODNAIF: Higher Order DNA Impact Factor

HODT: Higher Order Delaunay Triangles

HOI: Higher-Order Information

HO k -Means: Higher Order k -Means

HOPC: Higher Order Parallel Coordinate

HOR: Higher Order Radar

HORC: Higher Order Radar Chart

HOS: Higher Order Spider

HOTA: Higher Order Trajectory Area

HOTrM: Higher Order Treemaps

HOVD: Higher Order Voronoi Diagram

kNN: k^{th} Nearest Neighbour

kNNQ: k^{th} Nearest Neighbour Query

kNVD: k^{th} Nearest Voronoi Diagram

kOR: k^{th} Order Region

kORQ: k^{th} Order Region Query

Mul-HOI: Multivariate Higher-Order Information

OF: Order Factor

OIF: Order Impact Factor

OkVD: Order k^{th} Voronoi Diagram

OOKVD: Ordered Order k^{th} Voronoi Diagram

OVD: Ordinary Voronoi Diagram

PCMM: Parallel Coordinate Multivariate Measure

POIs: Points of Interest

RMS: Root Mean Square

SSE: Sum Squared Error

TF: Trajectory Factor

TIF: Trajectory Impact Factor

VA: Visual Analytics

VDM: Visual Data Mining

Vis: Visualisation

Table of contents

List of figures	xxi
List of tables	xxix
1 Introduction	1
1.1 Background	1
1.2 Motivations	5
1.2.1 Multivariate Data Support	5
1.2.2 Directional Information and Dissimilarity Measurement	7
1.2.3 Topological Information and Dissimilarity Measurement	8
1.2.4 Higher-Order Information Mining	9
1.2.5 Hierarchical Representative	11
1.3 Research Questions and Gaps	11
1.4 Research Aim and Contributions	14
1.5 Thesis Outline	18
1.6 Conclusioin	20
2 Preliminaries and Literature Review	21
2.1 Higher Order Information	21
2.1.1 Quantitative and Qualitative	22
2.1.2 Categories of Higher Order Information	22
2.1.3 k^{th} Nearest Neighbour Queries	23
2.1.4 k^{th} Order Region Queries	24

Table of contents

2.1.5	Data Structure of HOI	25
2.2	Higher Order Voronoi Diagram	27
2.3	Dynamicity	28
2.4	Univariate/Multivariate Dataset	29
2.5	Visual Analytics	31
2.5.1	Comparison of Visual Analytics/Visualization/Visual Data Mining .	32
2.5.2	Evaluation	33
2.6	Dissimilarity Measurement and k -means Clustering	33
2.7	Visual Analytics for Higher-Order Information	35
2.7.1	Basic Raster-based Approach	35
2.7.2	Basic Visualisation Approach	37
2.7.3	Basic Vector-based Approach	39
2.8	Conclusion	40
3	Research Framework	43
3.1	Underlying Data Structure	43
3.2	Input Trajectory Preprocessing	44
3.3	Visual Analytics Tools	45
3.4	Higher-Order Information Mining	47
4	Unified Data Structure and Input Trajectory Preprocessing	49
4.1	Unified Data Structure	49
4.2	Input Trajectory Preprocessing	53
4.2.1	Topologically Simplified	54
4.2.2	Topologically Interpolated	55
4.2.3	Discussion	56
5	Visual Analytics Techniques	57
5.1	Higher-Order Parallel Coordinate (HOPC)	57
5.1.1	Geometrical PC	59

5.1.2	Topological PC	60
5.1.3	Directional PC	61
5.1.4	Summary and Discussion	62
5.2	Higher-Order Spider (HOS)	62
5.2.1	Directional Spider Chart (DSC)	63
5.2.2	Directional Balance Ratio (DBR)	65
5.2.3	Directional Dissimilarity Measure (DDM)	68
5.2.4	Summary and Discussion	69
5.3	Higher-Order Radar (HOR)	70
5.3.1	Higher-Order Radar Chart (HORC)	71
5.3.2	Higher-Order k -Means (HOk-Means)	72
5.3.3	Summary and Discussion	74
5.4	Higher-Order Treemaps (HOTrM)	75
5.4.1	Binary Treemaps for HOI	77
5.4.2	Squarified Treemaps for HOIs	78
5.4.3	Strip Treemaps for HOIs	78
5.4.4	Slice Treemaps for HOIs	80
5.4.5	Ordered Treemaps for HOIs	81
5.4.6	Pivot by Size Treemaps for HOIs	82
5.4.7	Pivot by Split Size Treemaps for HOIs	83
5.4.8	Treemap Comparison	84
5.4.9	Summary and Discussion	86
5.5	Higher-Order DNA (HODNA)	87
5.5.1	Higher-Order DNA Chart (HODNAC)	88
5.5.2	Higher-Order Trajectory Area (HOTA)	89
5.5.3	Higher-Order DNA Impact Factor (HODNAIF)	90
5.5.4	Summary and Discussion	93
5.6	Conclusion	94

Table of contents

6	Evaluation	97
6.1	Efficiency	97
6.2	Comparison and Contrast	100
7	Emergency Management Application	103
7.1	HOPC for Emergency Management	104
7.1.1	Mitigation	105
7.1.2	Preparedness	106
7.1.3	Response	106
7.1.4	Recovery	107
7.1.5	Discussions	108
7.2	HOS Application	109
7.3	HOR Application	112
7.4	HOTrM Application	115
7.5	HODNA Application	117
7.6	Combination of HOPC and HOR for Emergency Management	120
7.6.1	Mitigation	121
7.6.2	Preparedness	122
7.6.3	Response	124
7.6.4	Recovery	125
7.7	Conclusion	129
8	Conclusion	131
8.1	Background	131
8.2	Summary of Contributions	133
8.3	Future Work	134
	References	137
	Appendix A Screenshots of proposed framework	147

List of figures

1.1	Multivariate POIs with trajectory dataset T_1 : (a) O1VD of generator $P_A = \{p_{A1}, \dots, p_{A10}\}$ (labelled from A1 to A10); (b) O1VD of generator $P_B = \{p_{B1}, \dots, p_{B10}\}$ (labelled from B1 to B10); (c) O1VD of generator $P_C = \{p_{C1}, \dots, p_{C10}\}$ (labelled from C1 to C10); (d) O1VD of 3 generators $P_A + P_B + P_C$	6
1.2	HOVD with 9 red generators $P_D = \{p_{D3}, \dots, p_{D11}\}$, and 3 trajectories $T_2 = \{t_{2.1}, \dots, t_{2.4}\}$ in green, $T_3 = \{t_{3.1}, \dots, t_{3.4}\}$ in blue, $T_4 = \{t_{4.1}, \dots, t_{4.4}\}$ in light-pink: (a) order-1 Voronoi diagram of P_D ; (b) order-2 Voronoi diagram of P_D	7
1.3	Voronoi diagrams with 13 generators $P_E = \{p_{E1}, \dots, p_{E13}\}$ (labelled from A1 to A13), and trajectory T_5 (12 timestamps): (a) order-1 Voronoi diagram of P_E ; (b) order-3 Voronoi diagram of P_E	9
1.4	Voronoi diagrams: (a) O1VD with 9 generators $P_F = \{p_{F1}, \dots, p_{F9}\}$ (labelled from A1 to A9), and 2 trajectories T_6 and T_7 (12 timestamps); (b) O1VD with 17 generators $P_G = \{p_{G1}, \dots, p_{G17}\}$ (labelled from A1 to A17), and 3 trajectories T_8 , T_9 and T_{10} (10 timestamps).	10
1.5	VA processes: (a) traditional VA; (b) proposed VA.	12
2.1	Raster-based vs. vector-based Voronoi diagram with 8 points: (a) bit-mapped (raster-based) graphics; (b) vector-based graphics.	27
2.2	Visual analytics processes.	31

List of figures

2.3	Raster-based weighted and obstacle Voronoi diagram for New Orleans case study.	36
2.4	Voronoi diagram of cushions and bevels visualisation approaches rendering for different metrics.	38
2.5	Vector-based Voronoi diagram families relationships.	39
3.1	Proposed research framework.	44
4.1	The OVD of $P_H = \{p_{H1}, \dots, p_{H6}\}$ (labelled from A1 to A6) and its dual Delaunay triangulation and circumcircle (a) OVD; (b) its dual Delaunay triangulation; (c) its corresponding circumcircle.	51
4.2	Multivariate POIs $P_I = \{p_{I1}, \dots, p_{I9}\}$ in red (labelled from A1 to A9), $P_J = \{p_{J1}, \dots, p_{J9}\}$ in blue (labelled from B1 to B9), and $P_K = \{p_{K1}, \dots, p_{K10}\}$ in purple (labelled from C1 to C10) with trajectory data set $T_{11} = \{t_{11.1}, \dots, t_{11.12}\}$ in green, $T_{12} = \{t_{12.1}, \dots, t_{12.12}\}$ in blue, $T_{13} = \{t_{13.1}, \dots, t_{13.12}\}$ in light-pink, $T_{14} = \{t_{14.1}, \dots, t_{14.12}\}$ in dark grey, and $T_{15} = \{t_{15.1}, \dots, t_{15.12}\}$ in brown: (a) Original dataset; (b) O1VD; (c) O3VD.	52
4.3	OkVD:(a) O1VD; (b) O2VD; (c) O3VD with trajectory $T_{16} = \{t_{16.1}, \dots, t_{16.10}\}$ and generator $P_L = \{p_{L.1}, \dots, p_{L.5}\}$ (labelled from A1 to A5); (d) topologically simplified T_{16} with O1VD; (e) topologically simplified T_{16} with O2VD; (f) topologically simplified T_{16} with O3VD.	54
4.4	OkVD:(a) O1VD; (b) O2VD; (c) O3VD with trajectory T_{16} and generators $P_M = \{p_{M.1}, \dots, p_{M.21}\}$ (labelled from A1 to A21); (d) topologically interpolated T_{16} with O1VD; (e) topologically interpolated T_{16} with O2VD; (f) topologically interpolated T_{16} with O3VD.	55
4.5	Examples of real datasets: (a) GPS datasets; (b) picture targets from Facebook.	56

5.1	OkVDs: (a) O1VD and (c) O5VD with generators $P_N = \{p_{N.1}, \dots, p_{N.8}\}$ in red (labelled from A1 to A8), $P_O = \{p_{O.1}, \dots, p_{O.8}\}$ in blue (labelled from B1 to B8), $P_P = \{p_{P.1}, \dots, p_{P.8}\}$ in magenta (labelled from C1 to C8) and trajectory data $T_{17} = \{t_{17.1}, \dots, t_{17.20}\}$ in green; (b) O1VD and (d) O5VD with generators, P_N for topologically simplified trajectory T_{17} in red, P_O for topologically simplified trajectory T_{17} in blue and P_P for topologically simplified trajectory T_{17} in magenta.	58
5.2	Geometrical PCMMs (x-axis: minimum (top) and maximum (bottom) of the distance to Order- k generators; y-axis: trajectory node ID): (a) Geometrical PCs - P_N ; (b) Geometrical PCs - P_O ; (c) Geometrical PCs - P_P based on the trajectory data T_{17} . Quantitative PCMM is shown in orange and qualitative PCMM is shown in cyan. When quantitative and qualitative PCMMs are the same, it is shown in green.	59
5.3	Topological PCMMs (x-axis: Order- k ; y-axis: trajectory node ID): (a) Topological PCs - P_N ; (b) Topological PCs - P_O ; (c) Topological PCs - P_P based on the trajectory data T_{17}	61
5.4	Directional PCMMs (x-axis: Order- k ; y-axis: trajectory node ID): (a) Directional PCs - P_N ; (b) Directional PCs - P_O ; (c) Directional PCs - P_P based on the trajectory data T_{17}	61
5.5	OkVDs with two trajectories $T_{18} = \{t_{18.1}, \dots, t_{18.101}\}$, $T_{19} = \{t_{19.1}, \dots, t_{19.101}\}$ and generators $P_Q = \{p_{Q.1}, \dots, p_{Q.13}\}$ (labelled from A1 to A13): (a) O1VD; (b) O3VD; (c) topologically simplified T_{18} and T_{19} with O1VD; (d) O3VD with two topologically simplified trajectory datasets T_{18} and T_{19}	64
5.6	Directional-geometrical DSCs: (a) O1VD DSC; (b) O3VD DSC; (c) O5VD DSC; (d) O10VD DSC with two trajectory datasets T_{18} , T_{19} and generators P_Q	65
5.7	Directional-topological DSCs: (a) O1VD DSC; (b) O3VD DSC; (c) O5VD DSC; (d) O10VD DSC with two topologically simplified trajectory datasets T_{18} , T_{19} and generators P_Q	66

List of figures

5.8	A well-balanced example: (a) O1VD; (b) DSC of O1VD based on directional-geometrical HOI DBR of trajectory datasets $T_{20} = \{t_{20.1}, \dots, t_{20.8}\}$ and generators $P_R = \{p_{R.1}, \dots, p_{R.6}\}$	67
5.9	Line graphs: (a) DBRs for directional-geometrical HOI; (b) DBRs for directional-topological HOI of trajectory datasets T_{18} and T_{19}	68
5.10	HORCs with the trajectory T_{17} used in Figure 5.1: (a) O1VD HORC based on P_N ; (b) O1VD HORC based on P_O ; (c) O1VD HORC based on P_P ; (d) O1VD HORC based on all P_N , P_O , and P_P together.	71
5.11	HORCs with the trajectory T_{17} used in Figure 5.1: (a) O5VD HORC based on P_N ; (b) O5VD HORC based on P_O ; (c) O5VD HORC based on P_P ; (d) O5VD HORC based on all P_N , P_O , and P_P together.	72
5.12	OkVDs: (a) O1VD; (b) O5VD; topologically simplified: (c) O1VD; (d) O5VD with trajectory datasets $T_{21} = \{t_{21.1}, \dots, t_{21.42}\}$ and $T_{22} = \{t_{22.1}, \dots, t_{22.62}\}$ and generators $P_S = \{p_{S.1}, \dots, p_{S.12}\}$ (labelled from 3 to 14).	76
5.13	Binary treemaps for HOI: (a) O1VD with topologically simplified T_{21} and T_{22} ; (b) O5VD with topologically simplified with T_{21} and T_{22} ; (c) O1VD with T_{21} and T_{22} ; (d) O5VD with T_{21} and T_{22}	77
5.14	Squarified treemaps of for HOI: (a) O1VD with topologically simplified T_{21} and T_{22} ; (b) O5VD with topologically simplified with T_{21} and T_{22} ; (c) O1VD with T_{21} and T_{22} ; (d) O5VD with T_{21} and T_{22}	79
5.15	Strip treemaps for HOI: (a) O1VD with topologically simplified T_{21} and T_{22} ; (b) O5VD with topologically simplified with T_{21} and T_{22} ; (c) O1VD with T_{21} and T_{22} ; (d) O5VD with T_{21} and T_{22}	80
5.16	Slice treemaps for HOI: (a) O1VD with topologically simplified T_{21} and T_{22} ; (b) O5VD with topologically simplified with T_{21} and T_{22} ; (c) O1VD with T_{21} and T_{22} ; (d) O5VD with T_{21} and T_{22}	81
5.17	Ordered treemaps for HOI: (a) O1VD with topologically simplified T_{21} and T_{22} ; (b) O5VD with topologically simplified with T_{21} and T_{22} ; (c) O1VD with T_{21} and T_{22} ; (d) O5VD with T_{21} and T_{22}	82

5.18	Pivot by size treemaps for HOI: (a) O1VD with topologically simplified T_{21} and T_{22} ; (b) O5VD with topologically simplified with T_{21} and T_{22} ; (c) O1VD with T_{21} and T_{22} ; (d) O5VD with T_{21} and T_{22}	83
5.19	Pivot by split size treemap for HOI: (a) O1VD with topologically simplified T_{21} and T_{22} ; (b) O5VD with topologically simplified with T_{21} and T_{22} ; (c) O1VD with T_{21} and T_{22} ; (d) O5VD with T_{21} and T_{22}	84
5.20	OkVDs: (a) O1VD and (b) HODNAC for trajectory $T_{23} = \{t_{23.1}, \dots, t_{23.13}\}$ based on $P_T = \{p_{T.1}, \dots, p_{T.13}\}$ (labelled from A1 to A13); (c) O1VD and (d) HODNAC for trajectory $T_{24} = \{t_{24.1}, \dots, t_{24.16}\}$ based on $P_T = \{p_{T.1}, \dots, p_{T.13}\}$ (labelled from A1 to A13).	89
5.21	HOTA: (a) HOTA for trajectory T_{23} based on P_T . (b) HOTA for trajectory T_{24} based on P_T	90
6.1	Comparison of existing HOI approaches: (a) Raster-based for O1VD; (b) Raster-based for O1VD with a line; (c) HOVD families with k NN and k OR; (d) Cushion and circle-region approach for Voronoi diagram; (e) Bevel approach for Voronoi diagram.	98
7.1	A study region and real world dataset: geo-tagged photos from Flickr. . . .	104
7.2	A study region and real dataset: distribution of clusters in the Central Business District (CBD) of Cairns (P_U in red), and an information office in the CBD of Cairns (P_V in blue), and an emergency office in the CBD of Cairns (P_W in magenta).	104
7.3	(a) O1VD; (b) O5VD trajectory $T_{25} = \{t_{25.1}, \dots, t_{25.50}\}$ with Mul-HOI $P_U = \{p_{U.1}, \dots, p_{U.14}\}$ (labelled from A1 to A14), $P_V = \{p_{V.1}, \dots, p_{V.12}\}$ (labelled from B1 to B12), $P_W = \{p_{W.1}, \dots, p_{W.10}\}$ (labelled from C1 to C10).	105
7.4	OkVDs: (a) O1VD; (b) O5VD with 3 new topologically simplified trajectories dataset. T_U in red, T_V in blue, T_W in magenta.	106
7.5	PCBCs: (a) generator P_U 's PCBC; (b) generator P_V 's PCBC; (c) generator P_W 's PCBC.	107

List of figures

7.6	Interaction PCBCs: (a) generator P_U 's PCBC; (b) generator P_V 's PCBC; (c) generator P_W 's PCBC.	108
7.7	A study region and real dataset: distribution of clusters in CBD of Cairns P_U	109
7.8	$OkVD$ with trajectories: (a) $O1VD$; (b) $O5VD$ with three trajectory datasets; (c) $O1VD$ with topologically simplified trajectories; (d) $O5VD$ with topologically simplified trajectories. Trajectory datasets T_{26} , T_{27} , and T_{28} with generators P_U	110
7.9	DSCs: (a) $O1VD$ directional-geometrical HOI; (b) $O5VD$ Directional-geometrical HOI; (c) $O1VD$ Directional-topological HOI; (d) $O5VD$ Directional-topological HOI with datasets shown in Figure 7.8.	111
7.10	Line graphs: (a) Directional-geometrical; (b) Directional-topological DBRs for datasets shown in Figure 7.8.	111
7.11	Line graphs: (a) Directional-geometrical; (b) Directional-topological DDMs for datasets shown in Figure 7.8.	112
7.12	(a) $O1VD$; (b) $O5VD$ with two trajectory datasets; (c) HORC of $O1VD$ (d) HORC of $O5VD$. Trajectory datasets T_{29} and T_{30} with generators P_U	113
7.13	(a) Original elbow method for $O1VD$; (b) $HO1$ -Means of trajectory datasets T_{29} and T_{30} with generators P_U	114
7.14	(a) Original elbow method for $O5VD$; (b) $HO5$ -Means of trajectory datasets T_{29} and T_{30} with generators P_U	114
7.15	$OkVDs$ with two trajectories T_{31} , T_{32} and a set P_U of generators: (a) $O1VD$; (b) $O5VD$ with two trajectory datasets; (c) $O1VD$ with topologically simplified trajectories; (d) $O5VD$ with topologically simplified trajectories.	115
7.16	Geometrical HOI and topologically simplified HOI treemaps with datasets shown in Figure 7.15: (a) squarified treemap with $O1VD$; (b) squarified treemap of topologically simplified datasets with $O5VD$; (c) pivot by split size treemap with $O1VD$; (d) pivot by split size treemap of topologically simplified datasets with $O5VD$	116

7.17	HOVD and HODNAC: (a) O1VD; (b) O5VD with generator P_U and trajectory datasets T_{33} in green, T_{34} in blue and T_{35} in light-pink; (c) HODNAC for trajectory T_{33} ; (d) HODNAC for trajectory T_{34} ; (e) HODNAC for trajectory T_{35} ; (f) HOTA with trajectory area T_{33} in green, T_{34} in blue and T_{35} in light-pink.	119
7.18	HOVDs for trajectory T_{36} : (a) O1VD and (b) O5VD of trajectory T_{36} with Mul-HOI P_U, P_V, P_W ; (c) O1VD and (d) O5VD with 3 new topologically simplified trajectories. T_{36U} in red, T_{36V} in blue, whilst T_{36W} in magenta. . .	121
7.19	Geometrical PCMMs: (a) Generator P_U 's geometrical PCMM for T_{36U} ; (b) Generator P_V 's geometrical PCMM for T_{36V} ; (c) Generator P_W 's geometrical PCMM for T_{36W}	122
7.20	Topological PCMMs: (a) Generator P_U 's topological PCMM for T_{36U} ; (b) Generator P_V 's topological PCMM for T_{36V} ; (c) Generator P_W 's topological PCMM for T_{36W}	123
7.21	Directional PCMMs: (a) Generator P_U 's directional PCMM for T_{36U} ; (b) Generator P_V 's directional PCMM for T_{36V} ; (c) Generator P_W 's directional PCMM for T_{36W}	123
7.22	Geometrical PCMMs: (a) Generator P_U 's geometrical PCMM for T_{36U} ; (b) Generator P_V 's geometrical PCMM for T_{36V} ; (c) Generator P_W 's geometrical PCMM for T_{36W}	124
7.23	Topological PCMMs: (a) Generator P_U 's topological PCMM for T_{36U} ; (b) generator P_V 's topological PCMM for T_{36V} ; (c) Generator P_W 's topological PCMM for T_{36W}	124
7.24	Directional PCMMs: (a) Generator P_U 's directional PCMM for T_{36U} ; (b) Generator P_V 's directional PCMM for T_{36V} ; (c) Generator P_W 's directional PCMM for T_{36W}	125
7.25	HORCs for O1VD: (a) Generator P_U 's HORC for T_{36U} ; (b) Generator P_V 's HORC for T_{36V} ; (c) Generator P_W 's HORC for T_{36W} ; (d) Generator P_U, P_V , and P_W 's HORC for T_{36}	126

List of figures

7.26	HORCs for O5VD: (a) Generator P_U 's HORC for T_{36U} ; (b) Generator P_V 's HORC for T_{36V} ; (c) Generator P_W 's HORC for T_{36W} ; (d) Generator P_U, P_V , and P_W 's HORC for T_{36}	127
7.27	Distributions of SSE for HOI (x-axis: order k ; y-axis: SSE) (a) Original elbow method for O1VD; (b) HO1-Means for O1VD; (c) Original elbow method for O5VD; (d) HO5-Means of trajectory dataset T_3 with generators P_U, P_V and P_W	128
7.28	Emergency management research framework.	129
A.1	Screenshots of proposed framework for HOPC.	148
A.2	Screenshots of proposed framework for HOTA.	149
A.3	Screenshots of proposed framework HOS.	150
A.4	Screenshots of proposed framework for HOR.	151

List of tables

1.1	Summary of research gap (Uni/Mul: Univariate POI/Multivariate POI; Quan/Qual: Quantitative HOI/Qualitative HOI; Geo: Geometrical; Top: Topological; Dir: Directional).	13
1.2	Research contributions (Geo: Geometrical, Top: Topological, Dir: Directional).	16
2.1	Various types of higher order information.	23
2.2	Summary of k^{th} nearest neighbour queries.	24
2.3	Summary of k^{th} order region queries.	25
2.4	Comparison and contrast of visual analytics/visualization/visual data mining.	32
2.5	Summary of visual anlaytics for higher-order information (Uni/Mul: Univariate PoI/Multivariate PoI; Quan/Qual: Quantitative HoI/Qualitative HoI; Geo: Geometrical; Top: Topological; Dir: Directional).	40
4.1	The relationship between HOVD families and the vector-based unified data structure.	50
5.1	Research contribution for HOPC.	62
5.2	Research contribution for HOS.	70
5.3	Research contribution for HOR.	75
5.4	Comparison of treemaps in terms of aspect ratio and distance change. . . .	85
5.5	Research contribution for HOTrM.	87
5.6	Research contribution for HODNA.	94
5.7	Summary of proposed VA tools.	94

List of tables

6.1	Effectiveness and efficiency comparison with existing approaches (N : the number of grids; n : the number of POIs. Typically N is much larger than n).	99
6.2	Comparison and contrast of proposed VA tools.	101
7.1	Comparison with the Flickr example dataset.	118

Chapter 1

Introduction

1.1 Background

We are living in a more convenient but also more complex and competitive environment, and society has moved into a more data-rich, information-rich but knowledge-poor environment. Due to the advances in technologies and the explosion of data, uncertainty in decision-making in this modern society has become an important issue [1]. People can gain access to a rich set of data through mobile devices and the web anytime, anywhere [2]. At the same time, much dynamic data is generated when people or objects are moving around. Sources of these data are coming from many different entities and devices. These created a dynamic environment where locations of the entities described by the data are moving and their relationship with other entities in the surrounding area is also changing. Now, the question is how to find the required patterns and knowledge for decision-making in this dynamic and uncertain environment, efficiently and effectively. To deal with these dynamic and uncertain situations, holistic planning, backup plans, and alternatives are necessary in case the original plan fails to work as intended. “What-if” analysis is the study of how the output of an analysis model varies with changes to inputs [3]. It is particularly useful for decision-making when uncertainties are present. For instance, what is the second nearest hospital from my current location when the first nearest hospital is closed or full? What if the first to third nearest hospitals are unavailable? Where is the fourth nearest hospital, how far and which direction

Introduction

is it from my current location? What are the patterns of the third nearest hospital in the study region?

This kind of Higher-Order Information (HOI) provides decision-makers quick alternative solutions whenever the first order information is not functioning properly. The basic data to capture movement of people and events are referred to as trajectory data. Two most widely requested HOI types relating to trajectory data are k^{th} Nearest Neighbour (kNN) information and k^{th} Order Region (kOR) information [4]. Given a set of n objects and a location P , kNN , analysis returns a set of k^{th} nearest objects from P based on a certain distance metric, while kOR yields a region to which P is the k^{th} closest. kNN has been widely used for classification, regression, and “what-if” analysis, and kOR has been extensively used for location optimisation, catchment analysis, market analysis, disaster management, and also “what-if” analysis [3]. Based on the characteristics of HOI, HOI can be classified into three different types: geometrical HOI, topological HOI, and directional HOI. Existing research of HOI mainly focuses on geometrical HOI, while limited attention has been paid to topological HOI and nothing to directional HOI. Geometrical HOI provides the location, space, and size information of kNN and kOR . However, relationship (topological) and positioning (directional) information of kNN and kOR will be strongly requested for “what-if” analysis. There are no integrated tools allowing users to investigate all these three types of HOIs.

The widespread use of low-cost Global Positioning System (GPS) equipped mobile devices has added another dimension to HOI. With this extra dimension, spatial and temporal data can capture dynamic details about the movements of people and events which are of interest in many different application domains [5, 6]. Take tourism as an example: tourists are interested in finding out what are the three nearest tourist attractions while they are driving or travelling around, or tour guides are interested in finding out where are the nearest spots most tourist groups have visited in the last 10 hours. Tour companies can organise a new business plan based on a region with good feedback. This dynamic HOI is also very important in other application domains, such as farmers who want to find out the four nearest ear-tagged unhealthy cows while drive-feeding, and policy makers who wish to find out the the furthest region for a new nursing home while visiting crime hotspots. Although they

may visit the same city, different driving or travelling paths can have different sight-seeing spots and views. Trajectories captured from moving entities in the geographical space are valuable sources of information. A trajectory can be represented as a set $P = p_1, p_2, \dots, p_n$ of continuous points, where each data point consists of a geospatial coordinate set and a time stamp such as $p_i = (x_i, y_i, t_i)$ for $p_i \in P$. Trajectory datasets contain not only information about the spatial and temporal aspects of moving entities but also additional quantitative and qualitative attributes related to the moving entities where the movement takes place.

While existing research has proposed different ways to analyse HOI for trajectory analysis [7, 8], existing methods can only deal with univariate Points of Interest (POIs). As our world moves into a multiple dataset support environment, there are many different type of datasets available to support decision-making. To analyse multiple datasets along with dynamic data, a multivariate HOI (Mul-HOI) analysis (HOI analysis for dynamically moving objects with respect to multivariate PoIs) approach is needed. Hence, it is important to understand the multivariate trajectory HOI (movements of HOI) of moving objects. There is a strong need for a technology solution that can efficiently and effectively help users to explore and understand this multivariate trajectory HOI. Despite the importance of moving objects HOI, analysis of trajectory HOI has received little attention mainly due to the complexity of moving objects' HOI [9].

Visualisation is a popular exploratory tool for people to manipulate and comprehend complex datasets. For instance, a basic approach to represent movement is to use arrows and flow lines to visualise time-varying changes [10]. To understand HOI such as the k NN and k OR information, some research has investigated different approaches and analytical tools for visualisation. Palmer [11] investigates visual aspects of HOI (highlighting unordered k OR) through contour lines and textures based on raster-based structure. Another approach [12] has attempted to visualise HOI regions using coloured edges and shading area. Basically, these approaches focus on visually deciding which are the k^{th} sites that influence a given partition of geometrical information. However, these approaches are not interactive and only provide very simple information to the users.

Introduction

Visual Analytics (VA) is a relatively new discipline that helps domain experts explore data by means of interactive interfaces and scientific visualisation. It is based on the fact that the human visual system is the best pattern detector [13]. VA is about representation, perception, data management, interaction, and decision-making. It utilises the strengths of human analytic capability and electronic data processing to enable humans to retrieve useful information from complex data. Recently, a few VA tools for trajectory data have been proposed [11, 12, 4]. Therefore, it is a solid candidate to be integrated with the complex dynamic HOI and gain insights with complex datasets [9, 1].

In addition to VA, quantitative analytics allowing users to quantify and compare HOIs are also important. Various techniques [14] have been proposed for processing, managing, and mining trajectory data in a broad range of application domains. Such application domains include surveillance analysis [15, 16], animal movement studies [17], traffic analysis [18, 19], sports performance analysis [20], supermarket shopping-paths optimisation [21], and emergency management [3, 22].

To deal with the complexity of trajectory datasets, researchers have investigated different approaches and analytical tools for mining trajectory datasets to reveal useful information and knowledge efficiently from this large amount of trajectory data [23]. For examples, there are approaches for clustering and analysing trajectory datasets, based on data mining techniques such as DBSCAN [24], k -means/ k -medoid [25], k -nearest neighbourhood [26], and regions-of-interest [27]. These approaches cluster trajectories into groups mainly based on optimising the similarity/dissimilarity measures. However, these dissimilarity measures are not able to deal with dynamic HOI.

To sum up, traditional approaches in existing research to analyse dynamic HOI share some common drawbacks: 1) a non-multivariate data structure model unsuitable for dynamic “what-if” analysis; 2) non-interactivity; 3) inability to support quantitative and qualitative types of information; 4) no-input data preprocessing techniques; 5) raster-based structure of HOI, thus incapable of visualising rich topological relations and directional positioning; 6) only for visualising a given k ; 7) inability to calculate dissimilarity measures based on geometrical, topological, and directional HOI; 8) no provision of HOI data mining approach;

9) lack of VA support. There is a strong need to develop new VA approaches to meet all these requirements.

This research investigates various VA approaches for dynamic Mul-HOI. First, we design a sound data structure that can effectively support dynamic Mul-HOI. The data structure not only supports the most widely used qualitative/quantitative geometrical HOI, but also topological and directional HOI. Second, we design and develop various VA tools that help users gain insights into the higher-order nature of data. Third, we introduce dissimilarity measurement approaches and data mining techniques to explore and recognise patterns. Finally we provide a cast study to illustrate the usefulness and applicability of the proposed methods.

1.2 Motivations

1.2.1 Multivariate Data Support

HOI for “what-if” analysis allows users to deal with the complexity of trajectory analysis. However, HOI for trajectory analysis has received little attention in the literature [28, 29]. HOI includes k NN information and k OR information that are of great significance when the first order or lower-order information is not functioning properly. For instance, the second nearest evacuation centre information is required when the first nearest evacuation centre is full or closed in emergencies. Analysing trajectory datasets and HOI could provide great benefits for emergency management.

While existing research has proposed different ways to analyse HOI for trajectory analysis [30], existing methods can only deal with univariate POIs. As our society moves into a more data-rich and information-rich but knowledge-poor environment, there are many different datasets available to support decision-making. To analyse multiple datasets along with the trajectory data, a Mul-HOI analysis approach is needed.

To compare the differences in univariate and Mul-HOI, consider Figure 1.1 with three sets of POIs data representing 10 tourist attractions (red points), 10 information centres (blue points), 10 emergency centres (magenta points), and 1 trajectory denoting tourist

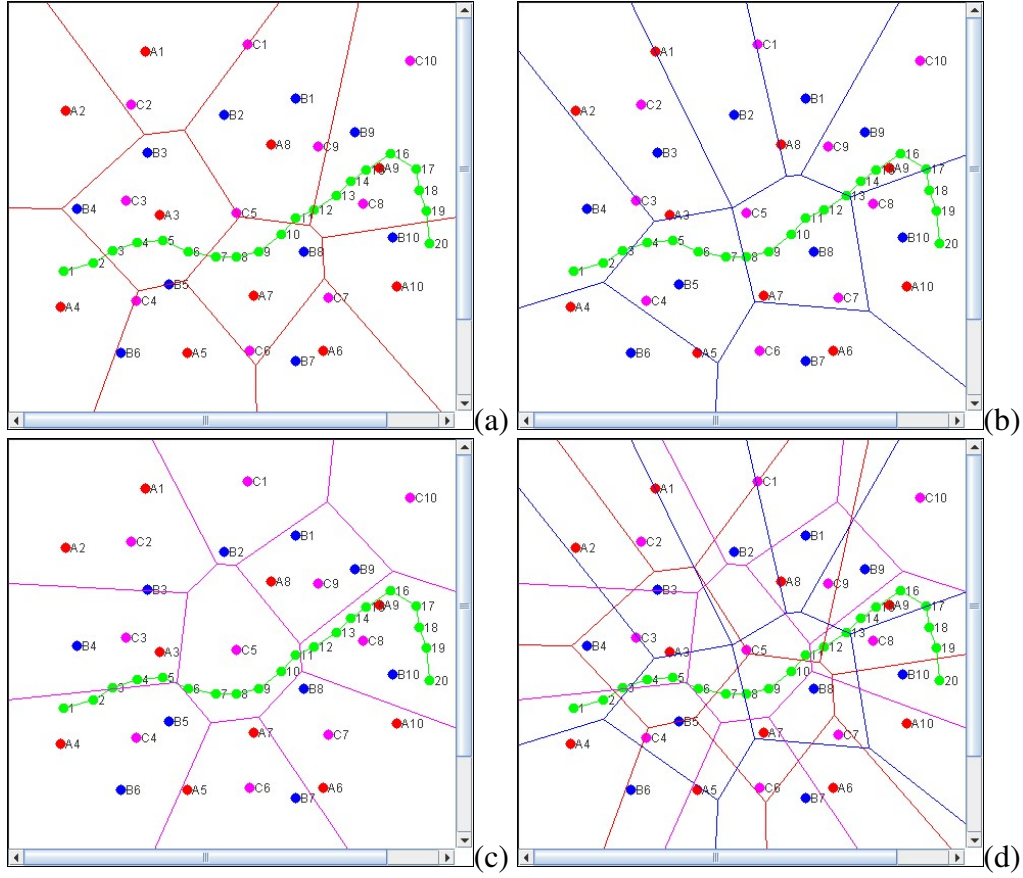


Fig. 1.1 Multivariate POIs with trajectory dataset T_1 : (a) O1VD of generator $P_A = \{p_{A1}, \dots, p_{A10}\}$ (labelled from A1 to A10); (b) O1VD of generator $P_B = \{p_{B1}, \dots, p_{B10}\}$ (labelled from B1 to B10); (c) O1VD of generator $P_C = \{p_{C1}, \dots, p_{C10}\}$ (labelled from C1 to C10); (d) O1VD of 3 generators $P_A + P_B + P_C$.

movements. The trajectory has 20 timestamps (nodes) representing spatial locations revealed from geo-tagged photos taken and uploaded to a photo-sharing website. From a univariate perspective, the geometrical HOI based on the nearest neighbours as the trajectory travels along the study region with regard to P_A ($p_{A4} - p_{A3} - p_{A7} - p_{A8} - p_{A9} - p_{A10}$) is shown in Figure 1.1(a). The trajectory geometrical HOI with regard to the generators P_B ($p_{B4} - p_{B5} - p_{B8} - p_{B9} - p_{B10}$) is shown in Figure 1.1(b). At the same time, Figure 1.1(c) shows the P_C sequence ($p_{C4} - p_{C3} - p_{C5} - p_{C8}$). In order to conduct multivariate analysis, we need to consider building HOI for the three sets of POIs together to retrieve additional information for complex multivariate analysis. Figure 1.1(d) shows geometrical HOI of those three sets

of POIs (generators) for the trajectory. This mixed sequence provides not only combined HOI for the three sets of POIs but also a way to build and analyse Mul-HOI. This Mul-HOI reveals interesting Mul-HOI patterns with trajectories not available with traditional univariate approaches. To analyse multiple datasets along with the trajectory data, a Mul-HOI analysis approach is needed.

1.2.2 Directional Information and Dissimilarity Measurement

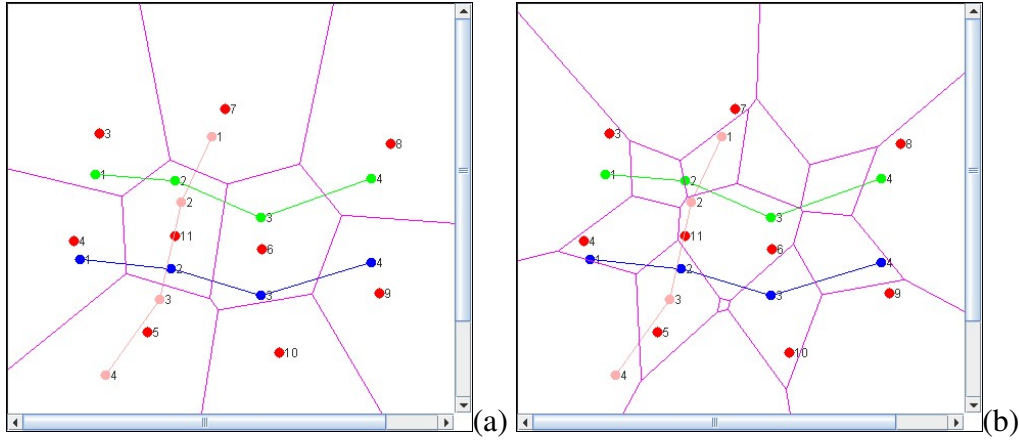


Fig. 1.2 HOVD with 9 red generators $P_D = \{p_{D3}, \dots, p_{D11}\}$, and 3 trajectories $T_2 = \{t_{2.1}, \dots, t_{2.4}\}$ in green, $T_3 = \{t_{3.1}, \dots, t_{3.4}\}$ in blue, $T_4 = \{t_{4.1}, \dots, t_{4.4}\}$ in light-pink: (a) order-1 Voronoi diagram of P_D ; (b) order-2 Voronoi diagram of P_D .

While most existing research has been related to HOI focusing on geometrical HOI, directional HOI for trajectory is of particular interest as very often, direction plays a critical role in decision-making. Trajectories exhibiting similar directional HOI also display similar directional dynamics. Consider Figure 1.2 with 9 generators P_D representing tourist attractions and 3 trajectories T_2 , T_3 and T_4 denoting tourist movements. Each trajectory has 4 timestamps (nodes) where geo-tagged photos were taken and uploaded to a photo-sharing website. T_2 and T_3 trajectories exhibit similar geometrical behaviours moving horizontally in the study area, moving from left to right (also the sequence of the first nearest tourist attraction is the same) while T_4 trajectory travels vertically in the study region from top to bottom. Most of the traditional trajectory analysis measures this kind of geometrical

Introduction

similarity and focuses on geometrical trajectory clustering, and regions-of-interests mining [31].

However, the T_2 and T_4 trajectories exhibit similar directional HOI with regard to tourist attractions. For instance, the T_2 trajectory ($t_{2.1}$ lies south of its nearest tourist attraction p_{D3} , $t_{2.2}$ lies north of p_{D11} , $t_{2.3}$ lies north of p_{D6} , and $t_{2.4}$ lies southwest of p_{D8}), exhibits the same directional HOI with the T_4 trajectory ($t_{4.1}$ lies south of p_{D7} , $t_{4.2}$ lies north of p_{D11} , $t_{4.3}$ lies north of p_{D5} and $t_{4.4}$ lies southwest of p_{D5}). That is, the T_2 and T_4 trajectories, despite having different geometrical information, exhibit the same directional sequences (south - north - north - southwest) with regard to its corresponding nearest neighbour tourist attractions in an order-1 Voronoi diagram (O1VD) as shown in Figure 1.2(a). Analogously, T_2 and T_4 trajectories show similar order-2 directional HOI in an order-2 Voronoi diagram (O2VD) as shown in Figure 1.2(b). All trajectory nodes (where they upload geo-tagged photos into the photo-sharing website) are located in between the two nearest neighbouring tourist attractions. This HOI (order-1 and order-2) reveals interesting similar directional dynamic patterns, which the traditional approaches fail to address, a directional HOI analysis approach is needed.

1.2.3 Topological Information and Dissimilarity Measurement

Consider an example shown in Figure 1.3, which shows different POI datasets and trajectory datasets with Voronoi diagrams. Figure 1.3(a) displays P_E , which contains 13 POIs data representing tourist attractions in O1VD labelled from A_1 to A_{13} (red points) and a trajectory T_5 denoting tourist movements with 12 time stamps (green points). The geometrical HOI is based on nearest neighbours as the trajectory travels along the study region with regard to the generators. In Figure 1.3(b) shows the corresponding order-3 Voronoi diagram (O3VD) of P_E and T_5 . For T_5 in Figure 1.3(a), the nodes can be classified into five clusters according to the tessellation of Voronoi diagram. In this case, the nodes of T_5 with timestamps (2-3-4) have their topological information changing each time, while for nodes with timestamps (8-9-10) the topological information is the same in O1VD. However, in Figure 1.3(b), nodes with timestamps (2-3-4) keep changing topological information as in the O1VD, and also nodes

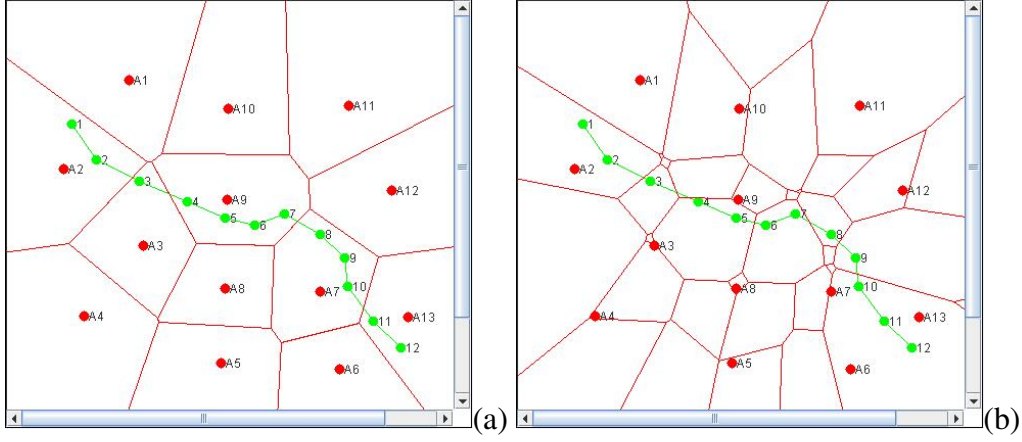


Fig. 1.3 Voronoi diagrams with 13 generators $P_E = \{p_{E1}, \dots, p_{E13}\}$ (labelled from A1 to A13), and trajectory T_5 (12 timestamps): (a) order-1 Voronoi diagram of P_E ; (b) order-3 Voronoi diagram of P_E .

with timestamps (8-9-10), topological information also changes each time. Hence, when only considering topological information in O1VD, this type of HOI related to the relationship between POIs and trajectory datasets will be ignored. There are several important questions to consider for decision-making. For example, which trajectory dataset is a more efficient way to move around the region or move along a required path? What is the best k -order information for the trajectory? How can we justify the impact and the relationship for a trajectory on each level of HOI, as well as the impact and relationship between each trajectory nodes. This example illustrates that a topological HOI analysis approach is needed.

1.2.4 Higher-Order Information Mining

In this era of “big data”, very often decision-makers are overwhelmed with the large amount of data made available to them. To find trends and similarities among different nodes in a large trajectory, one possible approach is to group similar nodes into clusters to reduce the complexity of the trajectory. Clustering in data mining is a process of grouping unlabelled objects into fewer classes of similar objects, called clusters. The main aim of clustering is to simplify the data, providing summarised information of the objects, and facilitating the discovery of hidden patterns [32]. Clustering is an important area of application for a variety

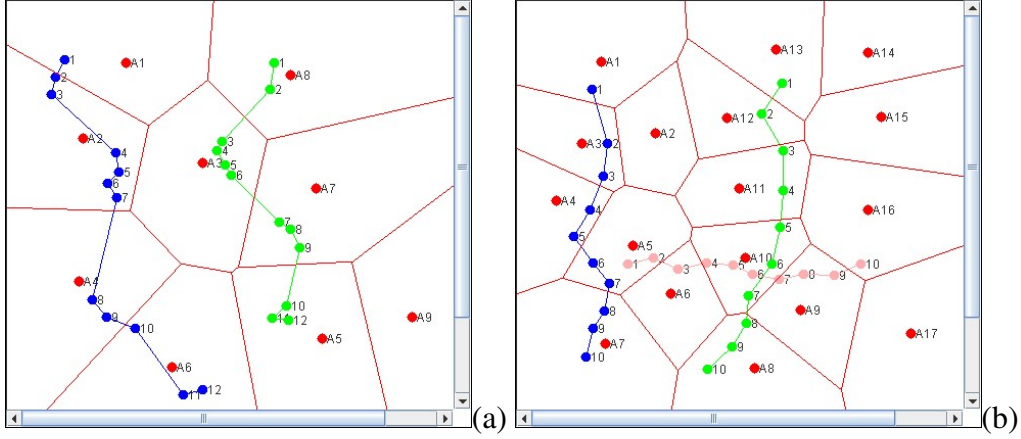


Fig. 1.4 Voronoi diagrams: (a) O1VD with 9 generators $P_F = \{p_{F1}, \dots, p_{F9}\}$ (labelled from A1 to A9), and 2 trajectories T_6 and T_7 (12 timestamps); (b) O1VD with 17 generators $P_G = \{p_{G1}, \dots, p_{G17}\}$ (labelled from A1 to A17), and 3 trajectories T_8 , T_9 and T_{10} (10 timestamps).

of fields including data mining, statistical data analysis, compression, and vector quantisation. There are different algorithms for clustering but among all of them, k -means clustering is popular for cluster analysis aiming to partition n observations into k clusters in which each observation belongs to a cluster with the nearest mean.

Consider an example shown in Figure 1.4 which shows different POIs datasets and trajectory datasets with Voronoi diagrams. Figure 1.4(a) displays P_F with 9 POIs labelled from A1 to A9 in red, and 2 trajectories T_6 and T_7 with 12 timestamps, T_6 is represented in green, T_7 is displayed in blue. In Figure 1.4(b) P_G shows 17 POIs labelled from A1 to A17 in red, and 3 trajectories T_8 , T_9 and T_{10} with 10 timestamps, T_8 is represented in green, T_9 is displayed in blue, T_{10} is presented in light-pink. Consider T_6 in Figure 1.4(a), using k -means clustering, the nodes can be classified into 4 clusters according to the tessellation of Voronoi diagram. Similarity, the nodes of T_7 can also be grouped into 4 clusters. However, in Figure 1.4(b), T_8 , T_9 , and T_{10} , each node in the trajectories is quite evenly separated from each other. When only considering the node distance, using k -means clustering, no meaningful clusters could be identified. Even though trajectory nodes can be defined over different groups based on existing k -means method, HOI is not considered during the clustering process. This demonstrates that a HOI mining approach is needed.

1.2.5 Hierarchical Representative

One of the challenges for analysing and visualising HOI of trajectory, is how to represent the large amount of information provided by HOI in a limited screen space. Treemaps are space-filling visualisation methods for representing large hierarchical collections of quantitative data [33]. They display hierarchical multi-level information as a set of rectangles and have been applied to a wide variety of domains including file systems structure [34, 35], financial analysis [36], decision-making processes [37], social cyberspace [38], and search-engine results [39]. Treemaps utilise a slice-and-dice approach to segment the space into a collection of rectangular areas. This hierarchical space segmentation is similar to vector-based space tessellations, such as Voronoi diagrams. A Voromap [40] is a space-filling approach that combines the Voronoi diagram and the treemap representation. It hierarchically represents Voronoi regions with multi-level Voronoi tessellations. However, it is for the ordinary Voronoi diagram, not for the higher-order Voronoi diagram. Therefore, a HOI treemaps approach is needed.

1.3 Research Questions and Gaps

Traditional VA consists of three main components: users, VA tools, and data. The user is a key and core element in VA. Different from information visualisation (typically for lay-persons), users in VA are often the domain experts. VA tools are not just for representing data in “visual means” to help users understand data easily but they are a series of interactive visual processes helping the users to gain insights about the dataset.

Typically, traditional VA involves large static data. Figure 1.5 depicts the traditional VA processes. Due to the nature of dynamic HOI, this traditional VA model could not be applied directly to our research for several reasons. First, data in the traditional VA model is typically static, while data in our research is dynamic. This dynamic research requires different data management and structure with which traditional data management approaches cannot cope. Second, the sizes of the data involved in the process are different. In traditional VA, data is large and VA tools are for complex and large data. That is, it is for data-rich

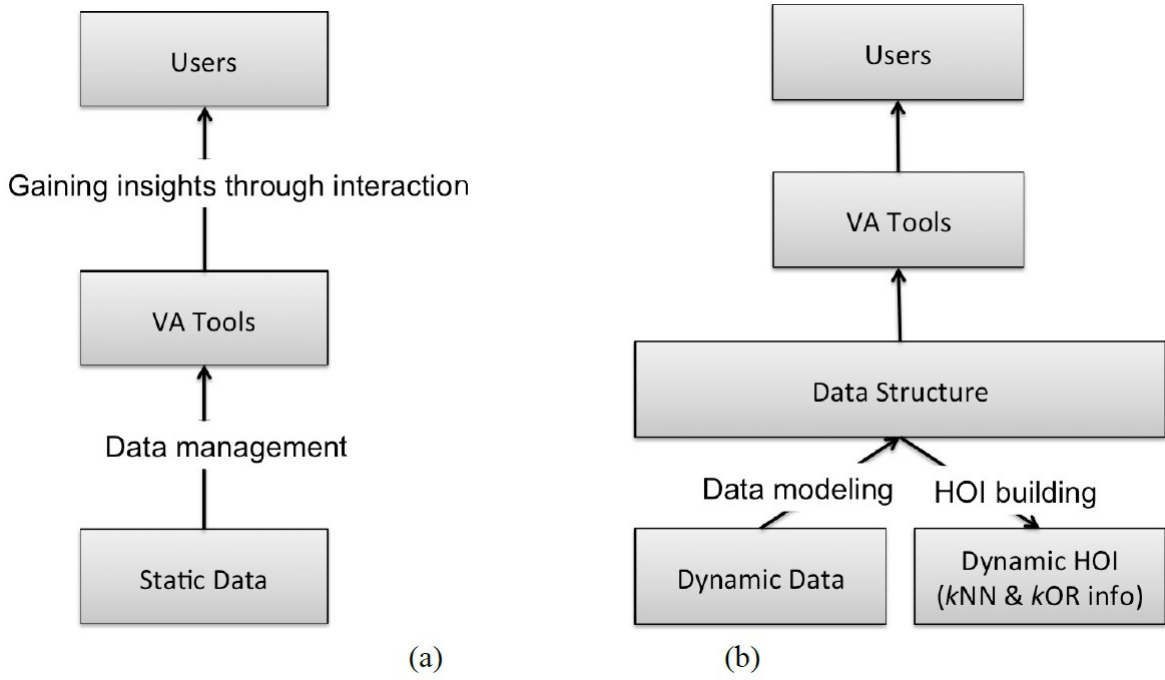


Fig. 1.5 VA processes: (a) traditional VA; (b) proposed VA.

but knowledge-poor situations. Data is not necessarily large in volume in our research, but HOI is. That is, kNN and kOR for all k values (1 to n) for quantitative/qualitative geometrical/topological/directional HOI is of large volume. Thus, it is for information-rich scenarios. Third, a data structure for a given data is not always required for traditional VA processes but VA tools are built on certain data management methods. Therefore, an efficient and effective data structure is a must in our research to support dynamic HOI. Fourth, the targets of visualisation are different. In a traditional VA model, the main visualisation target is data, while it is information (HOI) and patterns in our research. That is, traditional VA tools are for data-rich, but the proposed VA tools for our research are for HOI-rich.

Research questions in this research are as follows:

Question 1: What is the required data structure to deal with the characteristics of multivariate POIs targets?

Question 2: What is the required data structure for efficiently manipulating datasets with dynamically moving objects?

1.3 Research Questions and Gaps

Question 3: What are the VA tools required to analyse and visualise quantitative or qualitative geometrical HOI for moving objects?

Question 4: What are the VA tools required to analyse and visualise quantitative or qualitative topological HOI for moving objects?

Question 5: What are the VA tools required to analyse and visualise quantitative or qualitative directional HOI for moving objects?

Question 6: What are the metrics required for dissimilarity measure for dynamic geometrical, topological and directional HOI?

Question 7: What are the appropriate data mining algorithms to analyse the dissimilarity of trajectories based on the dissimilarity measure?

Table 1.1 Summary of research gap (Uni/Mul: Univariate POI/Multivariate POI; Quan/Qual: Quantitative HOI/Qualitative HOI; Geo: Geometrical; Top: Topological; Dir: Directional).

	Uni/ Mul	Dynamic Support	Inter-/ active	Quan/ Qual	Higher-Order Information (HOI)			Higher-Order Dissimilarity Measurement			Higher-Order Data Mining
					Geo	Top	Dir	Geo	Top	Dir	
Raster based approaches	Uni	No	No	Quan	Yes	Yes	No	No	No	No	No
k NNQ/ k ORQ approaches	Uni	No	No	Quan	Yes	Yes	No	No	No	No	No
Research gap	Mul	Yes	Yes	Quan/Qual	Yes	Yes	Yes	Yes	Yes	Yes	Yes

As a summary of the research questions, Table 1.1 shows the research gap in the literature. Previous research can be classified into two different groups, namely raster-based methods and k NNQ/ k ORQ methods. As Table 1.1 shows, previous research considers univariate data only, it cannot support trajectory datasets, it is non-interactive, it focuses only on quantitative information only, there is no higher-order dissimilarity measurement method, and is without higher-order or other data mining techniques provided. In terms of different types of HOI, although there are some methods for geometrical and topological HOI, there are no directional methods for HOI.

Thus, the research gap in this area is as follows:

Introduction

Univariate/Multivariate: no research for data structure supporting multivariate POIs;

Trajectory Support: no research for data structure supporting Mul-HOI (HOI for dynamically moving objects with respect to multivariate PoIs);

Interactivity: no interactive VA tools for the exploration of Mul-HOI;

Quantitative/Qualitative: no VA research considering both quantitative and qualitative Mul-HOI;

Higher-Order Information: some research for geometrical and topological HOI for static objects, but not for dynamically moving objects. In addition, no directional Mul-HOI;

Higher-Order Dissimilarity Measurement: no HOI dissimilarity measurement proposed;

Higher-Order Data Mining: no HOI mining approaches.

1.4 Research Aim and Contributions

The main aim of this research is to investigate VA tools to explore quantitative/qualitative geometrical, topological and directional HOI for moving objects with respect to multivariate POIs . Sub-aims of the research are as follows:

Aim 1: To propose a robust data structure supporting VA tools for multivariate POIs;

Aim 2: To extend the data structure to efficiently manipulate dynamically moving objects;

Aim 3: To develop VA tools for quantitative/qualitative geometrical HOI for moving objects with respect to multivariate datasets (geometrical Mul-HOI);

Aim 4: To develop VA tools for quantitative/qualitative topological HOI for moving objects with respect to multivariate datasets (topological Mul-HOI);

Aim 5: To develop VA tools for quantitative/qualitative directional HOI for moving objects with respect to multivariate datasets (directional Mul-HOI);

1.4 Research Aim and Contributions

Aim 6: To develop dissimilarity measures for geometrical/topological/directional Mul-HOI;

Aim 7: To develop Mul-HOI mining algorithms that return top- k similar trajectories based on the Mul-HOI dissimilarity measures.

Table 1.2 Research contributions (Geo: Geometrical, Top: Topological, Dir: Directional).

	Univariate/ Multivariate	Trajectory Support	Interactivity	Quantitative/ Qualitative	Higher-Order Information (HOI)			Higher-Order Dissimilarity Measurement			Higher-Order Data Mining	Research Aim
					Geo	Top	Dir	Geo	Top	Dir		
Research gap	?	?	?	?	?	?	?	?	?	?	?	-
Unified Data Structure (Chapter 4 - Section 4.1)	Mul	Yes	-	Quan/Qual	-	-	-	-	-	-	-	1
Topologically Simplified (Chapter 4 - Section 4.2.1)	Mul	Yes	Yes	Quan/Qual	Yes	Yes	No	No	No	No	No	2
Topologically Interpolated (Chapter 4 - Section 4.2.2)	Mul	Yes	Yes	Quan/Qual	Yes	Yes	No	No	No	No	No	2
Higher-Order Parallel Coordinates (Chapter 5 - Section 5.1)	Mul	Yes	Yes	Quan/Qual	Yes	Yes	Yes	Yes	No	No	No	3, 4, 5, 6
Higher-Order Spider (Chapter 5 - Section 5.2)	Mul	Yes	Yes	Qual	No	No	Yes	No	No	Yes	No	5, 6
Higher-Order Radar (Chapter 5 - Section 5.3)	Mul	Yes	Yes	Quan/Qual	Yes	Yes	Yes	No	No	No	Yes	3, 4, 5, 7
Higher-Order Treemaps (Chapter 5 - Section 5.4)	Mul	Yes	Yes	Quan/Qual	Yes	Yes	No	No	No	No	No	3, 4
Higher-Order DNA (Chapter 5 - Section 5.5)	Mul	Yes	Yes	Quan/Qual	No	Yes	No	No	Yes	No	No	4, 6

In this research, we propose a set of VA tools to explore and analyse Mul-HOI. The main contributions of this research are listed in Table 1.2 and details are as follows:

- to propose a unified vector-based data structure for supporting multivariate POIs targets, and for efficiently manipulating dynamically moving object datasets. This answers research aim 1, which is to introduce a robust data structure that supports multivariate POIs of VA tools.
- to propose data preprocessing approaches for supporting Mul-HOI. This achieves research aim 2, which is to extend the proposed data structure, and to effectively and efficiently manipulate dynamically moving objects.
- to present new VA tools based on parallel coordinates chart, spider chart, radar chart, treemap, and DNA chart for analysing large trajectory datasets with three categories (geometrical, topological and directional) of Mul-HOI.
- to present a new VA tool called a higher-order parallel coordinate chart for the determination of parallel coordinates multivariate measure to quantify both qualitative and quantitative Mul-HOI with regard to k NN and k OR information. This answers research aims 3, 4, 5 and 6, which are to develop VA tools for quantitative and qualitative, geometrical, topological and directional HOI for moving objects with respect to multivariate datasets; and to develop geometrical dissimilarity measures of Mul-HOI.
- to present a new VA tool called a higher-order spider chart for the determination of directional balance ratio and directional dissimilarity measure to quantify both qualitative and quantitative Mul-HOI with regard to k NN and k OR information. This answers research aims 5 and 6, which are to develop VA tools for quantitative and qualitative, directional HOI for moving objects with respect to multivariate datasets and to develop directional dissimilarity measures of Mul-HOI.
- to present a new VA tool called a higher-order radar chart for the determination of higher-order k -means clustering for trajectory data mining to quantify both qualitative and quantitative Mul-HOI with regard to k NN and k OR information. This answers

research aims 3, 4, 5 and 7, which are to develop VA tools for quantitative and qualitative, geometrical, topological and directional HOI for moving objects with respect to multivariate datasets and to develop Mul-HOI data mining algorithms that return top- k similar trajectories information.

- to present a new VA tool called a higher-order treemap for the determination of seven different types of treemap methods to quantify the best presentation method for both qualitative and quantitative Mul-HOI with regard to k NN and k OR information. This answers research aims 3 and 4, which are to develop VA tools for quantitative and qualitative, geometrical and topological HOI for moving objects with respect to multivariate datasets.
- to present a new VA tool called a higher-order DNA chart for the identification of a moving object's unique topological colour-code and the determination of higher-order trajectory area and higher-order DNA impact factor to quantify both qualitative and quantitative Mul-HOI with regard to k NN and k OR information. This answers research aims 4 and 6, which are to develop VA tools for quantitative and qualitative, topological HOI for moving objects with respect to multivariate datasets and to develop topological dissimilarity measure of Mul-HOI.
- to present a case study to demonstrate the usefulness and applicability of these new VA tools.

1.5 Thesis Outline

This dissertation was undertaken as a thesis by publication. The thesis is divided into eight chapters. Chapter 4 through Chapter 7 contain a content that has been either published or submitted for publication, in the form of three journal articles and five conference papers, all double-blind peer reviewed. The chapters are organised as follows:

Following this introduction, we first provide the necessary background information and related work that is required in order to understand the work performed in this thesis. The

background begins with an overview of HOI. First, we describe details of quantitative and qualitative HOI. Second, we discuss categories of HOI and summarise various types of HOI. Following this, we define the k NN queries and k OR queries and summarise their limitations. Next, we explain the details of HOI data structure. Following this, dynamicity and univariate/multivariate datasets are reviewed. We explain the definition, comparison and evaluation of VA. We describe dissimilarity measurement and k -means clustering and explain how to extend previous approaches for this research. At last, more detailed raster-based approaches, basic visualisation methodologies and vector-based approaches related in this research are presented.

Chapter 3 proposes an overall research framework and provides details of each main component. It also describes how each component is structured to answer research questions and research aims.

Chapter 4 introduces a vector-based unified data structure and two input trajectory preprocessing algorithms: topologically simplified and topologically interpolated. The former, topologically simplified preprocessing, is useful when an input trajectory is densely recorded whilst the latter, topologically interpolated preprocessing, is useful when the input trajectory is sparsely recorded.

Chapter 5 describes details of the five VA techniques which are higher-order parallel coordinate, higher-order spider, higher-order radar, higher-order treemaps and higher-order DNA, including geometrical, topological and directional dissimilarity measurement methodologies, and HOI mining approaches.

Chapter 6 evaluates our proposed framework in comparison with past studies. Also, this chapter compares proposed VA tools with respect to their abilities to handle multivariate POIs, trajectories, interactivity, and Mul-HOI.

Chapter 7 presents a real-world application, detailing how each VA algorithm analyses tourist datasets for emergency management. This chapter also compares the performance of each algorithm and provides the answer to research aims.

Chapter 8 concludes the findings of our research and summarises potential future work.

1.6 Conclusioin

Overall, this chapter provides background information on the general VA for dynamic HOI in Section 1.1, and motivations of this research with a strong need of VA tools for dynamic and multivariate HOI in Section 1.2. We identify research gaps and formulate research questions in Section 1.3, which leads to our aims and hypotheses in Section 1.4. Table 1.2 summarises our research aims and main contributions.

Chapter 2

Preliminaries and Literature Review

This chapter outlines related work that has contributed and led to the development of the techniques introduced in this thesis. This chapter starts with details of quantitative and qualitative HOI, and we discuss categories of HOI related queries and summarise their limitations. Next, we explain the details of HOI data structure. Following this, we review dynamicity and univariate/multivariate datasets. We explain the definition, comparison and evaluation of VA. We discuss dissimilarity measurement and k -mean clustering and explain how to extend existing approaches for this research. At last, more detailed raster-based approaches, basic visualisation and vector-based approaches related in this research are described.

2.1 Higher Order Information

Due to the complexity of modern society and advances in data collection devices, the source data is not static but dynamic, moving from place to place. “What-if” analysis is becoming more important than ever before in data analysis [3, 4]. When the first option is no longer available or not working anymore, a backup plan will receive attention and opportunity for people to use this information. It will be crucial to understand these complex and massive information sets, how to obtain useful patterns from those uncertain, changing over time, and how to filter and mine patterns hidden behind the large information.

HOI is information about more than one object for a given set of objects. That is, “Higher” means more than one target object and “Order” means the number of query (source) objects. Two most widely requested HOI types are k NN information and k OR information. In this research, HOI refers to k NN and k OR information. Note that, k NN utilises the term “Higher” while k OR employs “Order”.

“What-if” analysis is about how an output model varies with changes of different inputs. HOI is essential for “what-if” nearest queries, “what-if” neighbouring queries, “what-if” zoning queries, “what-if” facility location queries, and “what-if” routing queries [3]. Neighbourhood information is indispensable for many operations in spatial information systems with geo-data, and it needs to be maintained during the whole procedure of gathering and processing of geo-data.

2.1.1 Quantitative and Qualitative

HOI can either be quantitative or qualitative depending on the levels of measurement [41, 42]. The concept of qualitative methods, in some contexts, is equated with non-numerical procedures, and it is associated with the use of categories and classifications which can be objective or subjective. However, in other contexts, it could be understood as subjective interpretative methods for the study of the meanings and motives of actors. The concepts of quantitative methods are the operations of counting and measuring, the models based on the theory of functions and statistical procedures. In general, qualitative HOI is described by nominal (categorical) scale while quantitative HOI is labelled with different degrees such as ordinal, interval, or ratio scales.

2.1.2 Categories of Higher Order Information

HOI can also be divided into three categories depending on its geospatial characteristics [43]: geometrical, topological, and directional. Geometrical HOI is information about shape, size, location, and the properties of space. For quantitative methods, it concerns length, area, and volume with formal mathematical science. However, for qualitative methods, it considers

Table 2.1 Various types of higher order information.

	Geometrical [43]	Topological [43]	Directional [43]
Quantitative [41, 42]	1. 1 st , 2 nd , 3 rd ... NN coordinate information; 2. area of 1 st order region; 3. shape of 2 nd order region.	1. 1 st , 2 nd , 3 rd ... NN relationship information; 2. Space relationship of 1 st , 2 nd , 3 rd ... k^{th} order region.	1. 1 st , 2 nd , 3 rd ... NN position and direction information; 2. Directional relationship of 1 st , 2 nd , 3 rd ... k^{th} order region.
Qualitative [41, 42]	1. order- k (target, unordered k NN) and the rest; 2. order- k (source, unordered k^{th} generators) and the rest.	1. neighbouring regions of k OR; 2. Overlapping regions with k OR.	1. order- k (target, unordered k NN) positioning properties and the rest; 2. order- k (source, unordered k^{th} generators) positioning properties and the rest.

value k^{th} order information for calculating area and volume. Topological HOI analyses information of basic properties of relationships such as space, dimension, and transformation. Directional HOI is about relative positioning properties with distance information. Table 2.1 shows various types of HOI and examples of each category.

2.1.3 k^{th} Nearest Neighbour Queries

In recent years, k NN query (k NNQ) has been extensively studied in traditional spatial databases [3]. These types of queries are frequently used in Geographical Information Systems (GIS) and are defined as follows: given a set of spatial objects (or POIs), and a query point, find the k^{th} closest objects to the query point. An example of k NNQ application is a query to handle various scenarios in the four stages of emergency management (mitigation, preparedness, response, and recovery) [44]. The regular k NNQ has been extensively used and numerous algorithms have been proposed. The majority of the algorithms are aimed at n -dimensional objects in Euclidean spaces and are based on utilising one of the variations of multidimensional index structures [45, 46].

The formal definition of the k NNQ is presented as follows:

k NNQ Definition: Given a point set S , a query point q and a positive integer k , the k NNQ returns a point k NN set (k NNs) q which is a subset of S such that $d(q, r) \leq d(q, p)$ holds for any point $r \in k$ NNs(q) and for any point $p \in (S - k$ NNs(q)).

Table 2.2 Summary of k^{th} nearest neighbour queries.

Problems	Geometrical		Topological		Directional	
	Quantitative	Qualitative	Quantitative	Qualitative	Quantitative	Qualitative
$kNNQ$	[47–49, 28, 50–53, 29, 54]	[55–58]	[3]	[59, 60]	No	No

$kNNQ$ simply returns kNN , and typically returns quantitative geometrical (mostly) and topological information. It helps researchers find patterns or information directly in many areas such as climatology, chemistry, ecology, emergency management, epidemiology and physics as shown in Table 2.2.

The majority of the previous work on $kNNQ$ is based on computing the distance between a query and the objects. Although $kNNQ$ is for computing the distance between a query object q and an object o will automatically result in the computation of the distance between q and the objects that are closer to q than o . On the other hand, there remain many problems that have not been researched for experts to analyse $kNNQ$ as indicated in Table 2.2. $kNNQ$ has been studied in geometrical information using quantitative methods, but only a few researchers have considered topological information and no directional information. The problems of $kNNQ$ at this stage are as follows:

- most research focuses on quantitative information;
- most research focuses on geometrical information, only a few research on topological information [11, 61];
- no research on directional information.

2.1.4 k^{th} Order Region Queries

$kNNQ$ is supplemented by the kOR query ($kORQ$) which returns a region R for a certain subset $P_i^{(k)}$ in P where the set of kNN for any location l in R is the same as $P_i^{(k)}$. $kORQ$ can also be obtained by k^{th} -Nearest Point Diagram [4]. $kORQ$ associated with a given site is the set of points in the plane such that the site ranks number k in the ordering of the sites by distance from the point, and is a collection of one or more unconnected cells [62]. It is useful

Table 2.3 Summary of k^{th} order region queries.

Problems	Geometrical		Topological		Directional	
	Quantitative	Qualitative	Quantitative	Qualitative	Quantitative	Qualitative
$kORQ$	[47, 49, 28, 63, 52, 64, 29]	[55, 57, 65, 66]	[3]	[60, 54]	No	No

to establish a boundary on the number of cells, as their complexity can affect the accuracy of the final output – the k^{th} order Voronoi diagram.

It returns kNN Voronoi regions. Typically, it returns qualitative geometrical (mostly) and topological information. It is useful for researchers in emergency management, ecology, epidemiology, polymers, and physics as shown in Table 2.3.

However, the problems are limited studies have been conducted for experts to analyse $kORQ$ as can be seen from Table 2.3. $kORQ$ has been studied in geometrical information with qualitative methods but little research has been undertaken on topological information, and no research for directional information. The problems of $kORQ$ similar like $kNNQ$ are as follows:

- most research focuses on quantitative information;
- most research focuses on geometrical information, only a few research on topological information [11, 61];
- no research work on directional information.

2.1.5 Data Structure of HOI

Both kNN and kOR are modelled by Higher Order Voronoi Diagram (HOVD) families (order- k Voronoi diagram ($OkVD$), ordered order- k Voronoi diagram ($OOKVD$), k -nearest Voronoi diagram ($kNVD$)), that are generalisations of the Ordinary Voronoi Diagram (OVD). The OVD naturally divides the study region S into mutually exclusive and collective exhaustive sub-regions [62]. Thus, it is regarded as a point-to-region transformation. For a given set of $P = \{p_1, p_2, \dots, p_n\}$, the OVD $\mathcal{V}$ is a set of $\{V(p_1), \dots, V(p_n)\}$ Voronoi regions where each $V(p_i)$ (for $p_i \in P$) is the set of locations that are closer to p_i than other locations in S . The

Preliminaries and Literature Review

OkVD $\mathcal{V}^{(k)}$ is a set of all order- k Voronoi regions $\{V(P_1^{(k)}), \dots, V(P_l^{(k)})\}$, where the order- k Voronoi region $V(P_i^{(k)})$ ($V(P_i^{(k)}) \neq \emptyset$) for a certain subset $P_i^{(k)}$ consisting of k points out of P defined as:

$$V(P_i^{(k)}) = \{p | \arg \max_{p_r \in P_i^{(k)}} d(p, p_r) \leq \arg \min_{p_s \in P \setminus P_i^{(k)}} d(p, p_s)\}. \quad (2.1)$$

Note that the number of order- k Voronoi regions is $O(k(n-k))$ and the actual size of l is $O(N^2)$ (N is the size of grid) in $\mathbb{R}^2$ [67].

In the OkVD, k generators are not ordered, however, in some situations an ordered set (sequence) would be of interest. This is modelled by the OOkVD $\mathcal{V}^{<k>}$. It is a set of ordered order- k Voronoi regions $\{V(P_1^{<k>}), \dots, V(P_m^{<k>})\}$ ($V(P_i^{<k>}) \neq \emptyset$), where $m = n(n-1) \dots (n-k+1)$ and the ordered order- k Voronoi region of $P_i^{<k>} = \{p_{i1}, \dots, p_{ik}\}$ defined as:

$$V(P_i^{<k>}) = \{p | d(p, p_{i1}) \leq \dots \leq d(p, p_{ik}) \leq d(p, p_j), p_j \in P \setminus \{p_{i1}, \dots, p_{ik}\}\}. \quad (2.2)$$

$\mathcal{V}(P_i^{<k>})$ is a refinement of $\mathcal{V}(P_i^{(k)})$, namely $V(P_i^{(k)}) = \bigcup_{P_j^{<k>} \in A^{<k>}(P_i^{(k)})} V(P_j^{<k>})$, where $A^{<k>}(P_i^{(k)})$ is the sequence of all possible k -tuples made of $p_{i1}, \dots, p_{ik}$ [62].

One popular variant of the OVD similar to $\mathcal{V}^{(k)}$ is the k NVD. This is particularly useful when users are interested in only k^{th} nearest region. The k NVD $\mathcal{V}^{[k]}$ is a set of all k^{th} nearest Voronoi regions $\mathcal{V}^{[k]} = \{V^{[k]}(p_1), \dots, V^{[k]}(p_n)\}$, where the k^{th} nearest Voronoi region $V^{[k]}(p_i)$ is defined as:

$$V^{[k]}(p_i) = \{p | d(p, p_i) \leq d(p, p_j), p_j \in P \setminus \{k \text{ nearest generators of } p_i\}\}. \quad (2.3)$$

$\mathcal{V}(P_i^{<k>})$ is a refinement of $\mathcal{V}(P_i^{[k]})$. That is, $V(P_i^{[k]}) = \bigcup_{(p_{j1}, \dots, p_{jk-1}) \in A^{<k-1>}(P \setminus \{p_i\})} V((p_{j1}, \dots, p_{jk-1}, p_i))$, where $A^{<k-1>}(P \setminus \{p_i\})$ is the set of all possible $(k-1)$ -tuples consisting of $k-1$ elements out of $P \setminus \{p_i\}$ [68, 62].

2.2 Higher Order Voronoi Diagram

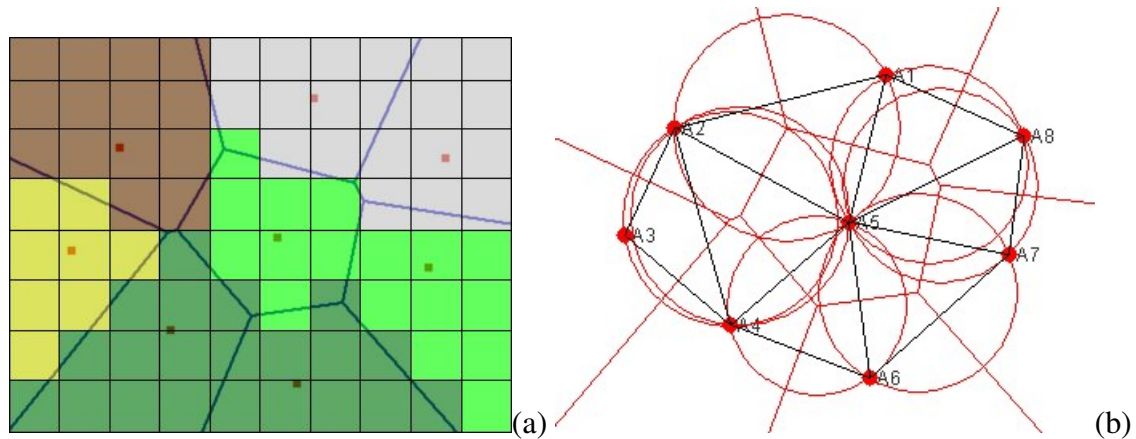


Fig. 2.1 Raster-based vs. vector-based Voronoi diagram with 8 points: (a) bit-mapped (raster-based) graphics; (b) vector-based graphics.

There are two different visual approaches that are able to derive Voronoi diagrams: one is raster-based as in Figure 2.1(a) and another is vector-based as shown in Figure 2.1(b). The comparison is as follows:

Raster-based HOVD: It is made up of pixels where each pixel is with a certain colour and arranged to display an image.

- pixel-based;
- dot matrix data structure;
- rectangular grid of pixels, points of colour;
- slow and poor quality.

Vector-based HOVD: It is made up of paths, each with a mathematical formula (vector) that tells the path how it is shaped and what colour it is bordered with or filled by.

- mathematical calculations of polygons;
- lead through locations called control points or nodes;
- fast and high quality.

The raster-based HOVD is able to construct approximate HOVDs with three popular Minkowski metrics (Euclidean, Manhattan and Maximum distance) and is able to visualise HOI information [69, 61]. However, this approach suffers from several major weaknesses that cannot be used for “what-if” analysis. First, it is raster-based without a data structure which is not suitable for Mul-HOI. Even it is not able to represent topological or directional HOI. Second, it needs to reconstruct HOVDs for every single task which is not time efficient for “what-if” analysis. Third, it does not compute exact HOVDs, but approximate HOVDs using a distance transformation. Therefore, this approach is not able to answer k NNQ and k ORQ precisely. Fourth, it is for a single static dataset and not suitable for multiple dynamic datasets. Fifth, it is a basic and simple visualisation tool and cannot support interactions with users.

The vector-based HOVD displays a complete set of HOVDs for k NNQ and k ORQ [3, 22]. It supports a user to explore different k values for various HOVDs. However, it is able to support geometrical HOI and basic topological HOI, but unable to support directional HOI. As in [3, 22], this approach only supports single static dataset and fails to deal with a moving object’s HOI analysis. In addition, this approach lacks high-level interactivity and HOI analysis to enable user-oriented queries.

2.3 Dynamicity

Dynamicity [70, 71] is about changes over time, and it is an important element in particular related to the spatio-temporal part of data. Time is not an additional dimension of spatio-temporal data but it needs to be handled as a special dimension of data as space is treated. Visualisation of dynamicity of moving objects is more about representing temporal aspects of the data. Moving objects are represented by trajectory datasets. There are several methods [72, 9] for temporal data as follows:

- linear vs. cyclic;
- time points vs. time intervals;

- ordered time *vs.* branching time.

Typical line graphs are used for general trends of linear time while spiral visualisations are useful for cyclic aspects of time data. Both linear and cyclic methods are of importance for dynamic HOI. Temporal data could be conceptually modelled as either time points or time intervals. Time points are instances without extent (duration) in time while time intervals are cases with extent. The former is more popular than the latter, but the latter is useful for representing imprecise or uncertain temporal primitives with duration. Time is also seen as an ordered collection of time such as from past to future, or branching of different courses of time. Both models are important aspects that can influence VA approaches in analysing data of trajectories. This research is about dynamic HOI of sequenced trajectory data on different time scales, thus this thesis will focus on linear time aspect, time points and order time aspect.

There are three basic types of spatial change with regard to moving objects [73] as below:

- change location in space (movement, changes);
- change relationship information of movement of objects;
- change in spatial extent by growing or shrinking and by changing shape.

This last change type is not directly relevant to this research since it is based on point data without extent. Thus, this research will investigate VA tools for the first two types of spatial changes. A review of visualisation use for temporal and spatial dynamicity of data can be found in [72, 9].

2.4 Univariate/Multivariate Dataset

Univariate and multivariate datasets [74] are the various types of data that are based on the number of variables. Variables mean the number of objects that are under consideration as a sample in an experiment. Usually there are two types of datasets as follows:

Univariate dataset: A univariate dataset is used for the simplest form of analysis. It is the type of data in which analysis is made based only on one variable. For example, there are 50 students in class year 4. If the variable marks obtained in English were the matter, then in that case, analysis will be based on the number of subjects that fall into defined categories of marks. In our thesis, one type of POI is an example of univariate dataset.

Multivariate dataset: A multivariate dataset is the data in which analysis is based on more than two variables per observation. Usually, multivariate data consists of a set of random variables that have multiple dimensions, is complex, and sometimes requires special software systems to aid in the analysis. In our thesis, a set of multiple POIs is an example of multivariate dataset.

While there are increasing efforts devoted to trajectory analysis, research considering trajectory HOI (HOI for dynamically moving objects) has received little attention in literature [28, 29]. Many different types of valuable information can be derived from trajectory data to understand the underlying behaviours or activities that generate the trajectory data and to help making smart decisions. For instances, k NN information and k OR information are the most popular research areas in HOI for trajectory data. HOI is of great importance in highly dynamic and complex environments. HOI for “what-if” analysis allows users to deal with the dynamicity and complexity of trajectory analysis especially when the best possible solution is unavailable/malfunctioning/fully booked. For instance, in emergency planning and management phases, HOI provides useful information to situations where more than k^{th} response centres are required to participate in the response and recovery phases.

The relationships between POIs and trajectories, and the comparisons of POIs and trajectories are getting more useful for different research areas. There is a need for tools that can find that information, and it is very important to understand that information from multivariate datasets for future research. In my research, the research framework is not only considering univariate dataset support but also supporting multivariate datasets for “what-if” analysis.

2.5 Visual Analytics

VA is a human-centric interactive visualisation-based analysis approach. It is based on the fact that the human's visual system is the best pattern detector [13]. With VA, humans solve complex problems with perceptual and cognitive capabilities. It also utilises the computing and processing power of computers, thus it combines the strengths of humans and computers [9, 1]. The main aim of VA is to enable analysts to use their perceptual and cognitive skills to gain insights, detect the expected, and discover the unexpected from data through progressively deepening analysis (overview first, zoom in and filter to show details-on-demand) [75], and steerable analysis through interactions [76].

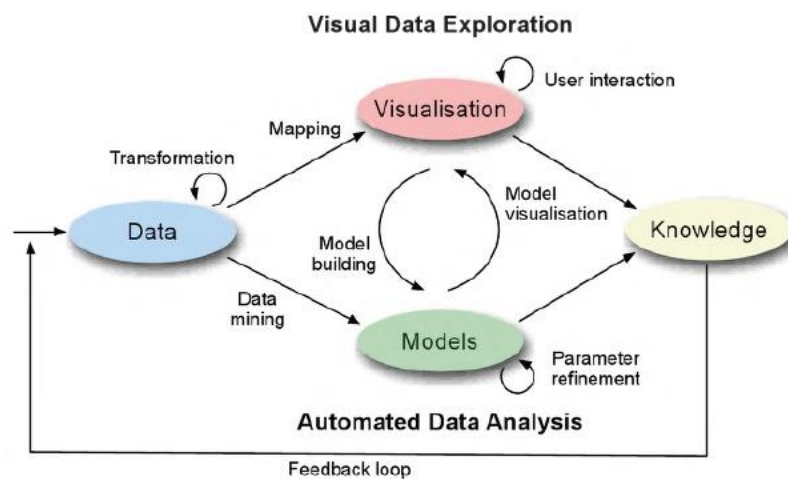


Fig. 2.2 Visual analytics processes.

VA processes are interactive and iterative. They are tightly coupled with human interactions. For an initial representation of VA, data analysis applies different statistical and mathematical techniques, such as spatio-temporal data analysis or mining depending on the dataset. The process then enters a loop where the user can get new insight knowledge from the dataset, ideally more focused and more adequate analytical techniques to drive the system. The user will understand visualisation and different views better, which helps him or her go ahead with the visual and eventually confirm hypotheses from previous iterations. Figure 2.2 depicts VA processes of data exploration [76].

Table 2.4 Comparison and contrast of visual analytics/visualization/visual data mining.

	Analysis Task	Decision Making	Interactive Visualisation	Data Mining Integration	Data Analysis	Focus	Data Size	Nature of Problem	Users
VA	Necessarily required	Integral approach to decision-making	Yes	Human-Centric	Exploratory data analysis	Data analysis through interactivity	Not necessarily large	Problem could be solved automatically	Experts
Vis	Not necessarily required	Not necessarily required	Not necessarily	Human-/Computer-Centric	Exploratory data Analysis	Producing visual views	Not necessarily large	Problems cannot be solved automatically	Everyone
VDM	Necessarily required	Limited approach to decision-making	Yes	Computer-centric	Exploratory data analysis, limited confirmatory data analysis	Data Mining through interactivity	Large	Problem could be solved automatically	Experts

2.5.1 Comparison of Visual Analytics/Visualization/Visual Data Mining

There is an overlay among “VA”, “Visualisation (Vis)” and “Visual Data Mining (VDM)”, but this subsection outlines main differences.

Table 2.4 compares and contrasts VA, Vis and VDM. Vis is Human/Computer-Centric, not necessarily requiring analysis, decision making or interaction with it. It is focusing on producing visual views for everyone, but problems cannot be solved automatically. It is for lay-people and experts while VA and VDM emphasise analysis functionality and they are designed for experts. VDM is Computer-Centric, requiring analysis and a limited approach to decision making through expert’s interactions. Problems could be solved automatically. VA is Human-Centric, it is an important distinction that requires analysis, and making decisions interactively with experts and it is not necessarily requiring large datasets from experts.

2.5.2 Evaluation

It is recommended that VA evaluated in terms of effectiveness, efficiency, and user satisfaction [76]. However, as in information Vis, evaluation in VA is a challenging task and there is no standardised approach for VA [76].

The main components of VA are users, data and VA tools. Effectiveness is a measure of recall (the fraction of relevant cases that are retrieved) and precision (the fraction of retrieved cases that are relevant), efficiency is a measure of computational time requirement and is a measure of meeting needs and expectations of users. That is, effectiveness measures whether VA tools meet their aims, efficiency assesses whether they produce timely responses, and user satisfaction calculates whether they are useful and intuitive.

Users are domain experts in VA. Typically, domain experts are limited in number, and they are busy with their professions, and it is hard to get domain experts in evaluation experiments. Often, non-experts could be tested on tools, but generalisation from lay-people to experts is hard. Also, subjectivity is another concern with users. VA is typically for ill-defined complex problems. There are unlikely known patterns and structures within the data. Evaluation of ill-defined problems is a daunting task, and there is no well-defined systematic evaluation approaches for ill-defined complex problems. Thus, due to the non-existence of ground truth for effectiveness evaluation and due to the very limited number of experts available, evaluations for effectiveness and user satisfaction are unrealistic to achieve for this study, instead a case study is used to demonstrate the usefulness and applicability of the proposed VA tools. Also this study includes time complexity analysis in order to prove the efficiency of the proposed framework.

2.6 Dissimilarity Measurement and k -means Clustering

Measuring dissimilarity is the heart of data analysis and data mining, in particular for algorithms used for discovering hidden patterns in data [77, 78]. To compare any two datasets or data properties, dissimilarity measurement allows users to handle numerical measures or qualitative reasoning to describe how different the two data objects are. Different

dissimilarity measures have been formulated in the literature, each with its own strengths and weaknesses [79, 80, 77]. However, existing dissimilarity measurement metric mainly focuses on geometrical type of information and they are not designed to deal with quantitative or qualitative geometrical, topological and directional trajectory HOI at the same time [79, 81]. To support informed decision-making for “what-if” analysis based on trajectory HOI, there is a need to have an effective dissimilarity measurement metrics for trajectory HOI to facilitate the comparison of HOI from multiple trajectories and multiple datasets. This thesis will investigate a new dissimilarity measure metric from existing traditional dissimilarity measures and modify them to handle complex trajectory HOI.

k -means clustering [82] is one of the most common clustering approaches. It is a method of vector-based quantisation. The goal of k -clustering is to partition the data points into k clusters with each data point to one of the k clusters by solving the optimisation problem which minimises the *distance* from the data point to the centroid of the cluster. The Euclidean distance is used in k -means clustering for the minimisation problem. This problem is NP-hard, a global minimum is not guaranteed. The proper number of clusters k to be used for clustering needs to be chosen by researchers with care. An incorrect choice of k will invalidate the clustering results. A numerical way to find the best k is to try k -means clustering over a range of k s and observing the resulting sum of squares for each k until it is converging to a local minimum [83]. Since k -means clustering is one of the most widely used data mining techniques and finding the best number of k is an important issue, this thesis will investigate an extension of k -means to handle complex traditional HOI to find Mul-HOI clusters, and an extension to find the best k using trajectory HOI. This research utilises OkVDs generated from the vector-based unified data structure and generates VA tools to determine the number of k for clustering with HOI. By extending the existing k -means clustering algorithm, this research defines a new algorithm that is able to handle clustering of HOI based on a new dissimilarity measurement proposed in this research.

2.7 Visual Analytics for Higher-Order Information

Even though k NNQ and k ORQ are useful for many applications, most research has been focused on developing algorithms or software applications for finding k NNQ and k ORQ [62]. VA aspects of k NNQ and k ORQ have received less attention. Furthermore, VA aspects of HOVD have received relatively poor attention due to their complex nature [11, 12, 4]. Interactive Vis can combine human interaction with computers to enhance focused visual illustrations of information. It is a good way to prune away unnecessary information and to focus on interested areas to spotlight. In the literature, relations of HOVD families and applications for the real-world problems have been studied [3] but no much research has been undertaken in HOI visualisation and VA aspects of HOI.

2.7.1 Basic Raster-based Approach

A list of algorithms to compute raster-based HOVDs has been proposed as discussed in Section 2.2 [84, 69, 61]. Some algorithms use Delaunay triangle-based data structures to compute raster-based HOVDs and others directly compute raster-based HOVDs without any data structure. The general idea of the raster-based method is to expand the Voronoi diagram incrementally by adding a point at a time for all points in the image by considering the number of neighbours or direction of connection [85–87, 62].

For example, [61] introduces a flexible sequential-scan based districting algorithm extended from the distance transformation algorithm. It extends the distance transformation algorithm to handle generators of complex primitives (point, line and area). New Orleans case study in Figure 2.3 shows that emergency services are put under review to try and understand why New Orleans suffered such a tragic loss during Hurricane Katrina. The Voronoi diagram is represented by a discrete grid lattice of size $N \times N$ which gives N^2 points in the plane to construct the complete Ok VD in $O(N^2)$ time [88]. The raster-based approach is a robust candidate for geospatial applications since it only supports geometrical (distance, coordinates) neighbourhoods information. One thing to note is that the raster-based approach

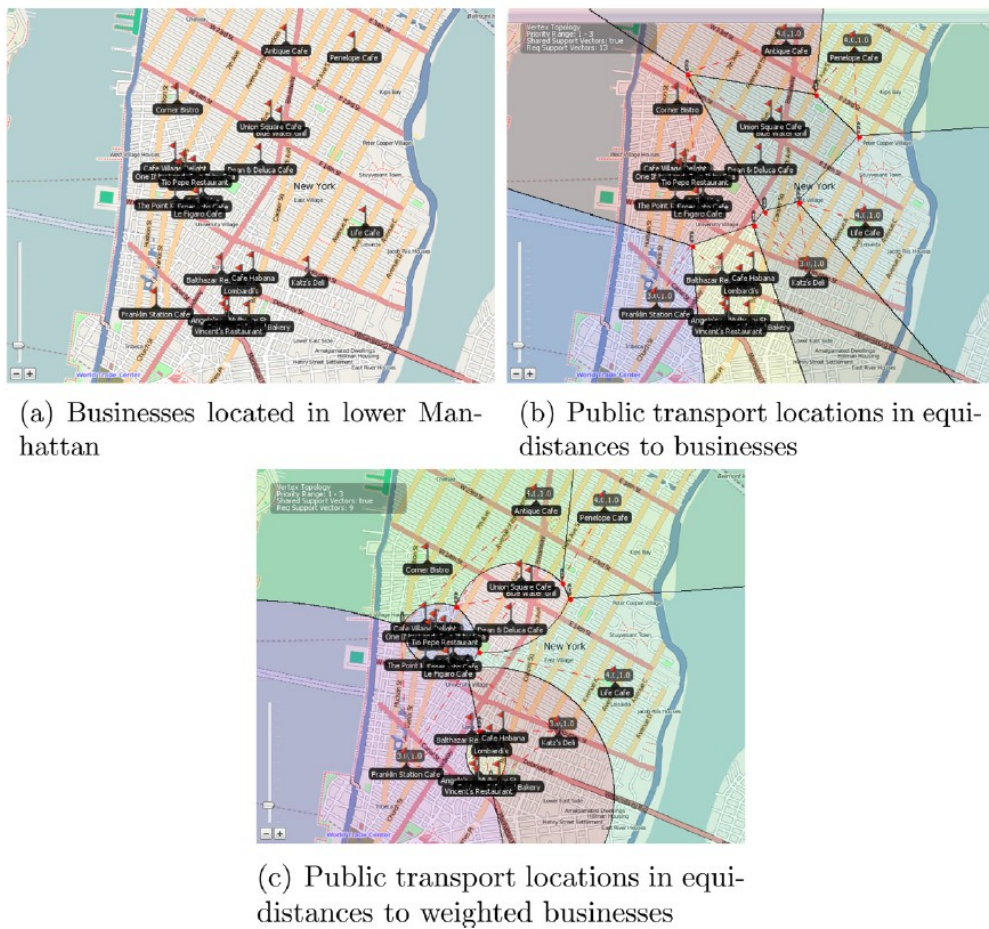


Fig. 2.3 Raster-based weighted and obstacle Voronoi diagram for New Orleans case study.

needs a very large N to generate approximate HOVDs which becomes computationally inefficient in data-rich environments.

2.7.2 Basic Visualisation Approach

Voronoi diagrams have been widely used for visualising domain-specific datasets [89–92], but little research has been conducted to visualise Voronoi diagrams themselves [11, 12]. The OVD has been typically visualised through Voronoi lines, vertices and regions [62]. Interactive visualisation could be used to visualise the OVD. Interactive visualisation is a visualisation approach that takes user input to focus on, and displays detailed information on focused area in real time [93]. It good to highlight areas of focus, but not good for producing an overall view simultaneously. Due to its filtering nature, interactive visualisation is particularly important in complex datasets or data-rich environments where scalability is an issue [94]. Voronoi growth models (similarly to contour models) could be used to visualise how each generator starts generating Voronoi regions [62]. When two growing regions meet each other, they stop growing and formulate a Voronoi edge. This growth model based visualisation dynamically shows how the OVD is formed, but it is not straightforward to use this approach to visualise HOVDs. This is because the OVD has one generator within its corresponding Voronoi region, but HOVDs have multiple generators not necessarily within a corresponding higher-order Voronoi region.

There has been some research on HOVD visualisation as depicted in Figure 2.4. Palmer [11] explores visual aspects of HOVD through contour lines and textures of raster-based HOVD. He introduces order- k plots (overlying nearest neighbour plots for $1, \dots, k-1$). That is order-3 plots are the O3VD overlaid with contour lines of order-1 and order-2 (1st nearest plot and 2nd nearest plot). However, it becomes more complex when k or n (n is the size of POIs) grows. Since it is raster-based Voronoi diagrams (without data structure model) and also does not support interactive visualisation, it is neither well suited for VA nor dynamic “what-if” analysis. Other researchers [12] attempt to visualise HOVD regions using colour and shading as follows:

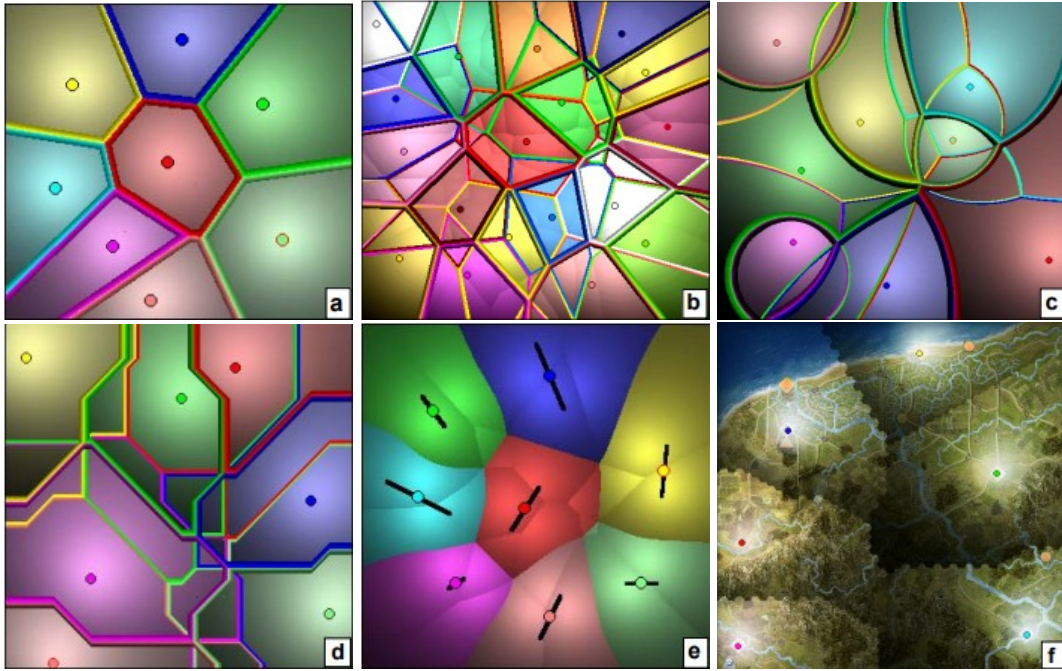


Fig. 2.4 Voronoi diagram of cushions and bevels visualisation approaches rendering for different metrics.

- Cushions: different intensity levels to give shading to represent the distance to the closest site. Brighter when closer to the generators while darker when close to the edges;
- Bevels: coloured for influencing sites and scaled to reflect the k^{th} order distance and hierarchy;
- K-Area/K-Edges: combination of Cushions and Bevels methods and extends to provide inter/intra/cross-relationships of HOVDs.

These three approaches are to superimpose k sites implicitly. [4] is using different widths and colours to highlight the relationship. It uses interactive visualisation to display the interested area. The method highlights topological relationships between/among HOVD families and enhances the readability of $OOkVD$.

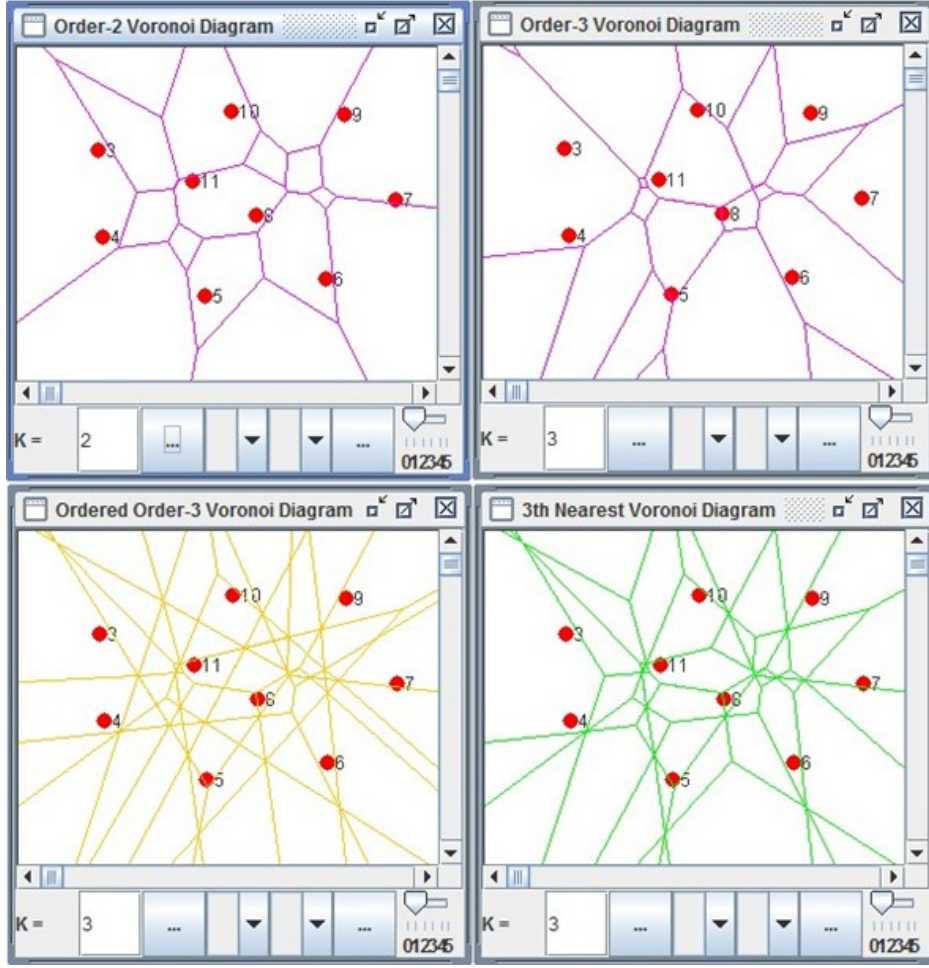


Fig. 2.5 Vector-based Voronoi diagram families relationships.

2.7.3 Basic Vector-based Approach

[3, 22] propose a unified Delaunay-triangle data structure that supports complete sets of all three HOVD families, shown in Figure 2.5, OkVD, OOkVD, and kNVD for all k^{th} order. Several other attempts [95, 96] have been made without data structure in order to improve efficiency. These approaches (without data structure) directly build OkVD from P . These without-data-structure-based approaches require less memory than the data-structure-based approaches but do not interactively support data analysis (unsuitable for VA). Building the data structure requires expensive time, and once it is determined, computing the complete set of HOVD families require $O(n^3 \log n)$.

2.8 Conclusion

HOI, dynamics and VA have been studied in many fields. Each of these is really popular and an important area in research topics. However, in Table 2.5, these traditional approaches share some common drawbacks: 1) non-multivariate data structure model is not suited for dynamic “what-if” analysis; 2) non-interactivity; 3) quantitative and qualitative type of information support; 4) non-input trajectory preprocessing techniques; 5) raster-based HOI data structure, thus incapable of visualising rich topological relations and directional positioning HOI; 6) only visualising a given order k ; 7) unable to calculate dissimilarity measure based on geometrical, topological, and directional HOI; 8) none of the HOI data mining approaches provided; 9) lack of VA techniques support. Thus, no research has been conducted with the combination of these three important topics. They are not able to answer the research questions we addressed in Chapter 1. This thesis will investigate this combination of three emerging areas to fill the research gaps. As Table 2.5 shows, this research will develop not only multivariate datasets supports, but also trajectory datasets supports. It will interactively communicate with users; focus both quantitative and qualitative information types; provide geometrical, topological and directional HOI; introduce dissimilarity measurement methods through 3 different HOI categories and HOI data mining techniques provided to return top- k trajectory information.

Table 2.5 Summary of visual analytics for higher-order information (Uni/Mul: Univariate PoI/Multivariate PoI; Quan/Qual: Quantitative HoI/Qualitative HoI; Geo: Geometrical; Top: Topological; Dir: Directional).

	Uni/ Mul	Dynamic Support	Inter-/ active	Quan/ Qual	Higher-Order Information (HOI)			Higher-Order Dissimilarity Measurement			Higher-Order Data Mining
					Geo	Top	Dir	Geo	Top	Dir	
Raster-based	Uni	No	No	Quan	Yes	Yes	No	No	No	No	No
Basic visualisation	Uni	No	No	Quan	Yes	Yes	No	No	No	No	No
Vector-based	Uni	No	No	Quan	Yes	Yes	No	No	No	No	No
Research gap	Mul	Yes	Yes	Quan/Qual	Yes	Yes	Yes	Yes	Yes	Yes	Yes

In conclusion, as reviewed in Section 2.1 HOI is of great importance in many dynamic scenarios. However, there have been limited studies in this area. Section 2.1.1 highlights

that both quantitative and qualitative HOI are useful and crucial but past studies mainly focused on quantitative HOI, but instead qualitative HOI has been ignored. Section 2.1.2 explains three types of HOI and their uses in different scenarios, however traditional studies mainly focus on geometrical HOI and leave topological and directional HOI unattended. Section 2.1.3 and Section 2.1.4 provide details and summaries of two popular research areas in HOI, k NNQ and k ORQ. Section 2.1.5 discusses the need of robust data structure for HOI manipulation and management, but there have been little studies in this field. Section 2.2 reviews HOVD and compares two different types of visual approaches. Section 2.3 discusses the data changing over time and summarises different types of moving objects. In addition, Section 2.4 highlights the need of multivariate data management whilst Section 2.5 identifies the need of effective VA for HOI manipulation. Section 2.6 reviews major developments in dissimilarity measures and shows a major gap in HOI similarity measure. Section 2.7 summarises past studies based on a thorough review analysis, and identifies main research gaps with regard to several important features. The research aims in Section 1.4 have been systematically developed to address these research gaps.

Chapter 3

Research Framework

The main aim of our research is to develop VA tools to provide various HOI for multivariate POIs to moving objects. Figure 3.1 depicts a detail of the proposed framework for this research. It shows the main components and their relationships to each other. The research questions can be answered by this research framework and targeted to research aims.

3.1 Underlying Data Structure

There is no existing VA tool to visualise HOI of moving objects (trajectory HOI) for multivariate POIs. In order to develop this VA tool, a solid data structure is required to support trajectory HOI for multivariate POIs. The unified data structure for HOVD [3] supports vector-based static HOI but does not efficiently support trajectory HOI for multivariate POIs. This research proposes a unified data structure based on Higher Order Delaunay Triangles (HODT), the dual of HOVD, and Higher Order CircumCircle (HOCC). These HODT and HOCC provide additional HOI that could be useful for multivariate POIs for k NN and k OR information. Previous research mainly focused on univariate data. However, multivariate data is more realistic, and supporting multivariate data is in great need. The unified data structure is a vector-based model supporting static HOI for univariate POI, and this thesis investigates extensions of this structure to trajectory HOI for multivariate POIs.

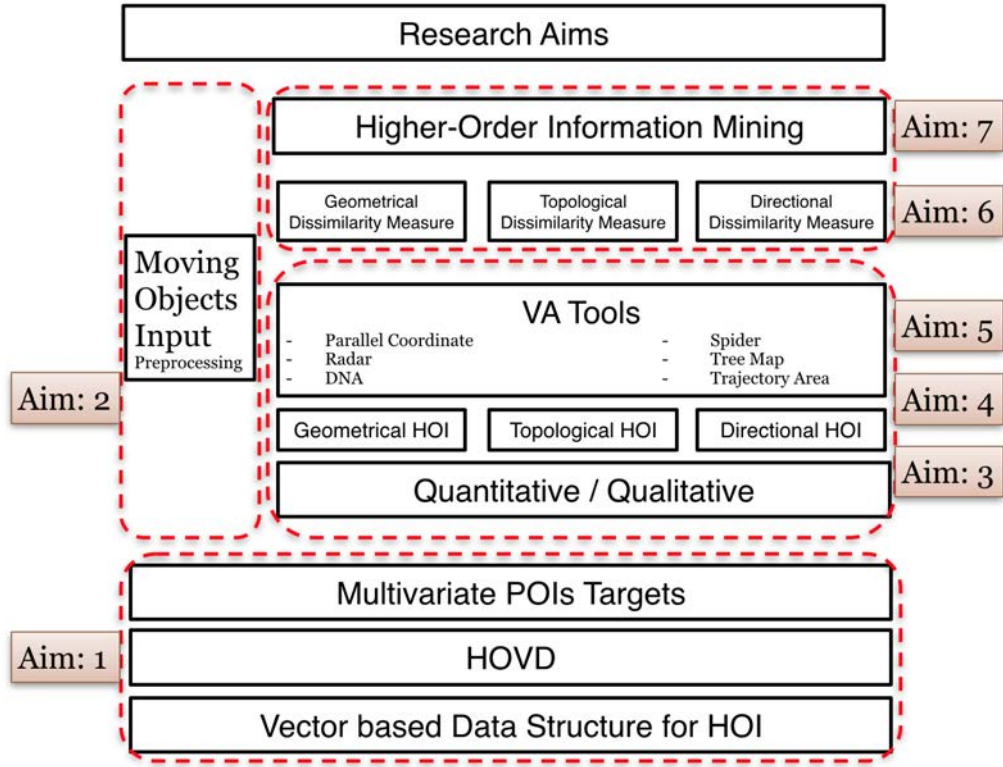


Fig. 3.1 Proposed research framework.

As shown in Figure 3.1, this corresponds to research question 1, which is, what is the required data structure to support the characteristics of multivariate POIs targets; and research aim 1, which introduces a robust data structure for supporting VA tools of multivariate POIs targets. Details are further discussed in Chapter 4.1.

3.2 Input Trajectory Preprocessing

Trajectory datasets of moving objects are very often collected with different types of mobile devices. These types of datasets contain noise, repeated data, and missing data uncontrollable by data users, and this creates many uncertainties in the decision-making process. There are many simplification approaches [97] to mix several ranking strategies and score heuristics together to create new simplification methods. However, existing ones are based on geometric

properties, but in this thesis topologically preprocessed approaches will be introduced and used.

This thesis introduces and innovates a new tool that can provide an optional method for filtering or grouping of moving objects data for users, which support research aim 2. Input datasets used in this thesis not only use raw/original trajectories, but also approach with topologically simplified trajectories when trajectories are too dense, and topologically interpolated trajectories when trajectories are too coarse.

As shown in Figure 3.1, this corresponds to research question 2, which is, what is the required data structure for dynamically moving objects with efficiently manipulating datasets; and research aim 2, which is to extend the data structure for providing dynamically moving objects efficiently. Details are further discussed in Chapter 4.2.

3.3 Visual Analytics Tools

Gaining insights is the main aim of VA applications. In many cases, insights are qualitative, and humans are more familiar with qualitative reasoning [3]. HOI could be categorised into two groups based on the nature of data: quantitative and qualitative. Efficient retrieval of both quantitative and qualitative information is a challenging task. This thesis investigates efficient approaches based on the vector-based unified data structure using HOVD. HOI can be divided into three clusters depending on the properties of geospatial characteristics: geometrical, topological, and directional information.

This thesis introduces and innovates efficient approaches that can be any combination of two categorisations (quantitative/qualitative category and geometrical/topological/directional category) to address research aim 3 to 5. The unified data structure could be used not only between geometrical, topological, and directional HOI, but also between quantitative and qualitative HOI from datasets of dynamically moving objects. There is a great deal of useful HOI hidden in these types of dynamic data. Current VA tools are designed for rich data but not for rich information. None of the existing VA tools is specifically designed for various dynamic HOI (quantitative/qualitative geometrical/topological/directional). Progressively

Research Framework

deepening the approach for efficiency, scalability, and steerable analysis will be utilised for maximal gaining and maximal insights. This thesis will investigate efficient and effective VA tools illustrating not only the quantitative and qualitative HOI but also the geometrical, topological, and directional HOI. A few new approaches will be introduced in Chapter 5. Each is briefly explained below:

Higher Order Parallel Coordinate presents a new visualisation approach by parallel coordinates for analysing large trajectory datasets with Mul-HOI and introduces a new measurement called parallel coordinates multivariate measure to quantify qualitative and quantitative Mul-HOI with regard to k NN information.

Higher Order Spider measures directional HOI for trajectories with regard to generators (targets). We introduce a directional spider chart-like visualisation approach and present a new measurement called directional balance ratio that quantifies directional HOI with regard to k NN information and another new measurement called directional dissimilarity measure.

Higher Order Radar introduces new geometrical and directional visualisation approaches and HOI k -Means clustering methods for analysing large trajectory datasets and HOI. These approaches are particularly efficient for providing new ways to represent geometrical HOI and directional HOI to provide semantic answers to problems such as top- k trajectory HOI.

Higher Order Treemaps are treemap representations of the HOVD. They present a robust way to model HOI through the HOVD and represent HOI for trajectory dataset through treemap approaches. Higher Order treemaps empirically implement well-studies seven different treemap layouts for trajectory dataset and evaluate their performance with two metric measures: aspect ratio and distance change ratio.

Higher Order DNA introduces new topological visualisation approaches, new geometrical visualisation approaches based on topological HOI, and HOI topological impact factor method for analysing the relationship between POIs and trajectory datasets.

This thesis will consider all three categories of HOI: geometrical HOI, topological HOI, and directional HOI. All three categories will support VA techniques to define and justify quantitative and qualitative HOI. As shown in Figure 3.1, this corresponds to research questions 3, 4, and 5, which are the required VA tools to analyse and visualise quantitative and qualitative, geometrical, topological and directional HOI of moving objects; and research aims 3, 4, and 5, which are, to develop VA tools containing quantitative and qualitative, geometrical, topological and directional HOI of moving objects with respect to multivariate datasets. Details will be further discussed in Chapter 5. The proposed VA tools can be used in different real-world applications to represent HOI as illustrated in Chapter 7.

3.4 Higher-Order Information Mining

As none of the existing research is specifically designed for various dynamic HOI (quantitative/qualitative geometrical/topological/directional) and data mining algorithms for trajectory HOI. With large amounts of trajectory data, it is important to obtain useful insights in a timely manner. Finding similarity and dissimilarity of different trajectories are ways to help the users digest large amounts of information. This thesis introduces and innovates different dissimilarity measurements that are based on geometrical, topological, and directional HOI based on proposed VA tools and these dissimilarity measurements will be able to find top- k similar trajectories.

As shown in Figure 3.1, this corresponds to research questions 6 and 7, which are, what are the metrics required for dissimilarity measurement approaches for geometrical, topological and directional HOI; and what are the appropriate data mining techniques to analyse the dissimilarity of moving objects based on dissimilarity measures; and research aims 6 and 7, which are, to develop dissimilarity measures for geometrical, topological and directional Mul-HOI; and to develop data mining techniques for Mul-HOI that return top- k similar moving objects based on dissimilarity measures. Details are further discussed in Chapter 5.

Chapter 4

Unified Data Structure and Input Trajectory Preprocessing

This chapter introduces a unified data structure of trajectory HOI for multivariate POIs, and new approaches of trajectory preprocessing for moving objects.

4.1 Unified Data Structure

The unified data structure of this thesis is vector graphics-based. In contrast to the traditional approach, which uses a raster-based data structure, this thesis proposes a new unified data structure based on vector graphics. Raster is made up of pixels, each with a certain colour, arranged in a dot matrix data structure to display an image. Compared with vector, raster is slow and of low quality. Vector is made up of paths, which can be represented by a mathematical expression (vector) that tells the path how it is shaped and what colour it is bordered with or filled with. Vector is mainly used in mathematical calculations of polygons with lead through locations called control points or nodes. Processing of vector is fast and able to produce high-resolution images without loss in quality.

A list of algorithms to compute HOVDs has been proposed in the literature [95, 96, 67]. Some algorithms use Delaunay triangle-based data structures to compute HOVDs, and others directly compute HOVDs without any data structure. The former better supports topological

Unified Data Structure and Input Trajectory Preprocessing

Table 4.1 The relationship between HOVD families and the vector-based unified data structure.

Order- k	$OkVD$	$OOkVD$	$kNVD$
1	Order-0 triangle Order-0 Circumcircle	O1VD	O1VD
2	Order-0 & Order-1 triangle Order-0 & Order-1 Circumcircle	O1VD-O2VD	O1VD-O2VD
3	Order-1 & Order-2 triangle Order-1 & Order-2 Circumcircle	O1VD-O3VD	O2VD-O3VD
4	Order-2 & Order-3 triangle Order-2 & Order-3 Circumcircle	O1VD-O4VD	O3VD-O4VD
...	...	...	...
k	Order- $(k-2)$ & Order- $(k-1)$ triangle Order- $(k-2)$ & Order- $(k-1)$ Circumcircle	O1VD- $OkVD$	O $(k-1)VD$ - $OkVD$

queries at the expense of memory. However, we focus on algorithms based on Delaunay triangle-based data structures since this research focuses on multivariate data structure and also multiple $kNNQ$ and $kORQ$.

The previous proposed algorithm is based on a Delaunay triangle-based data structure that constructs the complete $OkVDs$ in $O(n^3 \log n)$ time. This Delaunay triangle-based data structure is a robust candidate for geospatial applications since it supports both geometrical and topological neighbourhoods. However, it only supports the complete $OkVD$. [3] further studied this structure and extended it to the complete set of HOVD families ($\mathcal{V}^{(k)}$, $\mathcal{V}^{<k>}$ and $\mathcal{V}^{[k]}$). In the literature, inter-relationships of HOVD families and applications to real-world problems have been studied [3, 98].

The Delaunay triangulation is a dual graph of the OVD with edges connecting neighbouring points. It can be constructed by linking two adjacent Voronoi generators if they share a Voronoi edge together. The relationship between the two diagrams is shown in Figure 4.1. Accordingly, in Order-0 Delaunay triangles, no generator is on the circumcircle of each Delaunay triangle in the triangulation. However, there could be triangles whose circumcircles include a number of generators within them. Order-1 triangles are those triangles whose corresponding circumcircles include only one generator. Therefore, Order- k triangles are

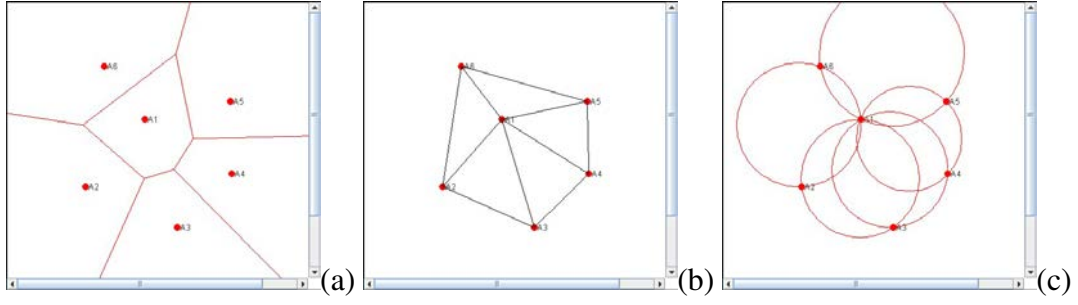


Fig. 4.1 The OVD of $P_H = \{p_{H1}, \dots, p_{H6}\}$ (labelled from A1 to A6) and its dual Delaunay triangulation and circumcircle (a) OVD; (b) its dual Delaunay triangulation; (c) its corresponding circumcircle.

those ones whose corresponding circumcircles include k generators. Please refer to [3] for more details.

A complete set of HOVD families could be computed from this set of Order- k Delaunay triangles [96, 3] and the relationship between the data structure and HOVD families is shown in Table 4.1. For order 1, O1VD, OO1VD and 1NVD are the same as the OVD that can be computed from the Order-0 Delaunay triangles (Delaunay triangulation). For order 2, O2VD is computed from both Order-0 Delaunay triangles and Order-1 Delaunay triangles ([96, 3] for detailed derivation). OO2VD is an overlay of O1VD and O2VD. 2NVD is the same as OO2VD. For order 3, O3VD is derived from both Order-1 Delaunay triangles and Order-2 Delaunay triangles. OO3VD is obtained by superimposing from O1VD to O3VD, whilst 3NVD is obtained by overlaying O2VD and O3VD. Similarly, Ok VD is calculated from both Order- $(k-2)$ Delaunay triangles and Order- $(k-1)$ Delaunay triangles. OOk VD is an overlay of all Ok VDs (from O1VD to Ok VD) whilst k NVD is an overlay of $Ok-1$ VD and Ok VD.

As our world moves into a multiple datasets support environment, univariate datasets have not provided enough information for experts to conduct an investigation into different research areas. Previous research areas provide data structures which can support the univariate datasets and provide the univariate HOI only. In this thesis, multiple datasets cannot be used with previous research data structures to display multiple targets HOI or moving objects HOI. Thus, our research aim 1 and aim 2 are to build a robust data structure which provides multiple POIs targets and to extend the data structure to efficiently manipulate moving objects.

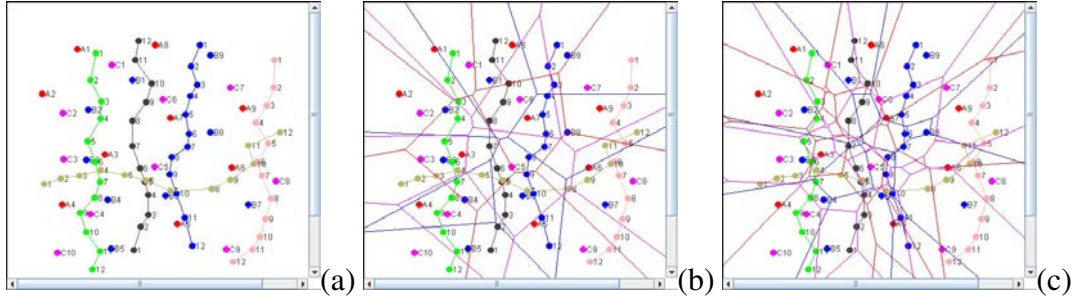


Fig. 4.2 Multivariate POIs $P_I = \{p_{I1}, \dots, p_{I9}\}$ in red (labelled from A1 to A9), $P_J = \{p_{J1}, \dots, p_{J9}\}$ in blue (labelled from B1 to B9), and $P_K = \{p_{K1}, \dots, p_{K10}\}$ in purple (labelled from C1 to C10) with trajectory data set $T_{11} = \{t_{11.1}, \dots, t_{11.12}\}$ in green, $T_{12} = \{t_{12.1}, \dots, t_{12.12}\}$ in blue, $T_{13} = \{t_{13.1}, \dots, t_{13.12}\}$ in light-pink, $T_{14} = \{t_{14.1}, \dots, t_{14.12}\}$ in dark grey, and $T_{15} = \{t_{15.1}, \dots, t_{15.12}\}$ in brown: (a) Original dataset; (b) O1VD; (c) O3VD.

Built on top of k NNQ and k ORQ information, we introduce a new approach to provide Mul-HOI for trajectory data, which is HOI analysis for dynamically moving objects with respect to multivariate POIs targets. The Mul-HOI implementation enables users to change various k values to obtain different types of Mul-HOI from the trajectory datasets. Users are able to interact with the program suite to retrieve different data points for any user-interested location within the set of query points [3, 22] on the fly. The Voronoi diagrams are drawn with a set of vector-based geometrical HOI. Geometrical HOI is computed based on the Euclidean distance of the trajectory data points to a set of nearest generators. This research utilises the unified Delaunay triangle based data structure which consists of a complete set of Order- k Delaunay triangles (from Order-0 to Order- $(k-1)$) [3] for supporting multivariate POIs and moving objects. Figure 4.2 displays screenshots of the proposed unified data structure with three sets of generators P_I in red, P_J in blue and P_K in purple, and trajectory datasets T_{11} in green, T_{12} in blue, T_{13} in light-pink, T_{14} in dark gray and T_{15} in brown. Figure 4.2(a) displays original datasets in different colours. Figure 4.2(b) presents O1VD and Figure 4.2(c) presents O3VD for those three sets of POIs for the trajectory in different colour (P_I Voronoi diagram in red, P_J Voronoi diagram in blue and P_K Voronoi diagram in purple). The details of VA approaches, how to analyse those hidden HOI will be explained in Chapter 5.

4.2 Input Trajectory Preprocessing

Most data preprocessing is based on geometric properties [97] and has been researched in many different areas. However, topology properties are receiving a great of attention in geo-information [22] and also in HOI [3]. This thesis introduces two new topology-based approaches to HOVD, namely simplification and interpolation.

The details of definition are as follows:

Topologically Simplified: when trajectories are too densely recorded or for sparse POIs, we use homology to measure the topological complexity of a point set. The simplest non-empty sets under this measure are the ones that contract to a point. Each such set consists of one component and simplifies based on the same k NN and k OR homological attributes. For example, when trajectory data has been densely recorded, datasets contain the same topological information, although geometrical information is different. In some cases, we have to consider topological attributes more than geometrical attributes. Topologically simplified can be used to filter trajectory datasets based on topological information.

Topologically Interpolated: when trajectories are too sparsely recorded or for dense POIs, we measure a new point based on the path of the trajectories through regions. If HOI regions contain empty data from the trajectory, then the mean of this empty region area is added into the trajectory datasets. Trajectories are interpolated based on the same k NN and k OR homological attributes. For example, when trajectory datasets have been sparsely recorded, they have provided geometrical information that can stop the trajectories sequential information based on topology attributes. Topologically interpolated can be used to build trajectory datasets sequential information based on topological information.

There are options for users to choose more suitable methods for input data analysis. In one case, when trajectory datasets are too densely recorded from GPS or mobile devices track locations, they are based on human capabilities such as domain knowledge perception and

creativity. In another case, when POIs are sparsely located from online data such as tourist locations or city emergency offices, trajectory datasets can be too dense for POIs. Trajectory datasets can be too sparsely recorded from Instagram or Flickr, which are online photo sharing websites. When POIs are densely located from restaurant or traffic lights locations, trajectory datasets can be too sparse for POIs. In such cases, topological interpolation could be in place. Considering time complexity, both topology property methods are $O(n \log m)$ when n is the number of source data, m is the size of POIs (targets). More details of topologically simplified and topologically interpolated HOI are shown in the next section. Methods to extend other VA methodologies are noted in Chapter 5. How to use those two methodologies related with a real-world study is presented in Chapter 7.

4.2.1 Topologically Simplified

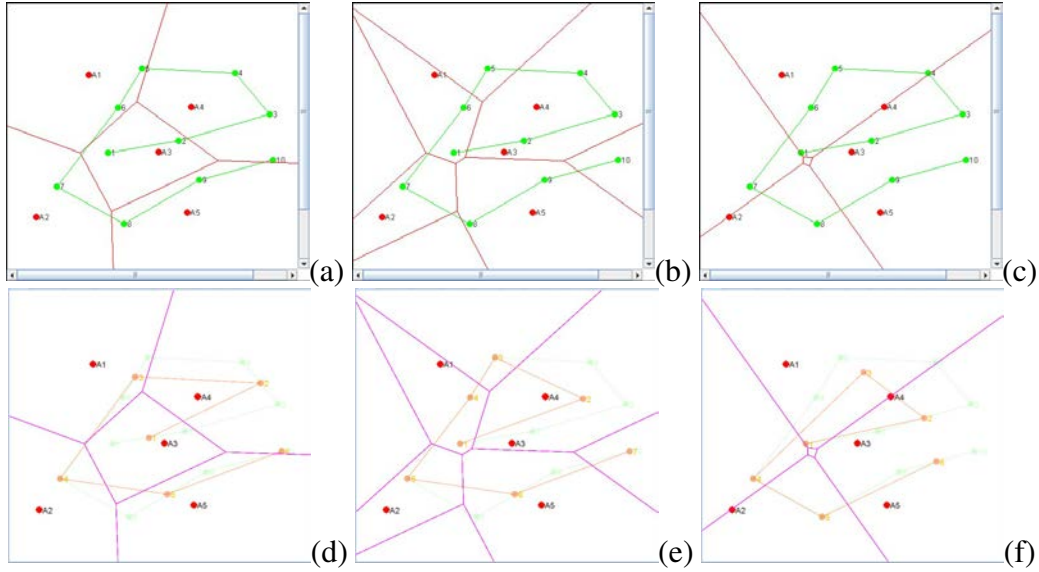


Fig. 4.3 OkVD:(a) O1VD; (b) O2VD; (c) O3VD with trajectory $T_{16} = \{t_{16.1}, \dots, t_{16.10}\}$ and generator $P_L = \{p_{L.1}, \dots, p_{L.5}\}$ (labelled from A1 to A5); (d) topologically simplified T_{16} with O1VD; (e) topologically simplified T_{16} with O2VD; (f) topologically simplified T_{16} with O3VD.

Figure 4.3(a), (b), and (c) show screenshots of visualisation tools with HOVD for a given POIs P_L and trajectory data T_{16} . Figure 4.3(a) shows O1VD, Figure 4.3(b) depicts O2VD and Figure 4.3(c) displays O3VD. Figure 4.3(d), (e), and (f) illustrate different topologically

simplified trajectory datasets based on different $OkVD$. For topologically simplified trajectory information, the visualisation tools will try to go through every single data point in all the data sets to perform a classification in relation to HOI. For data points, if they have the same k^{th} order information, then they are clustered into the same group. This approach is faster compared to other data mining approaches. Some examples of visualisation methods of topological simplified trajectory HOI are shown in the next chapter.

4.2.2 Topologically Interpolated

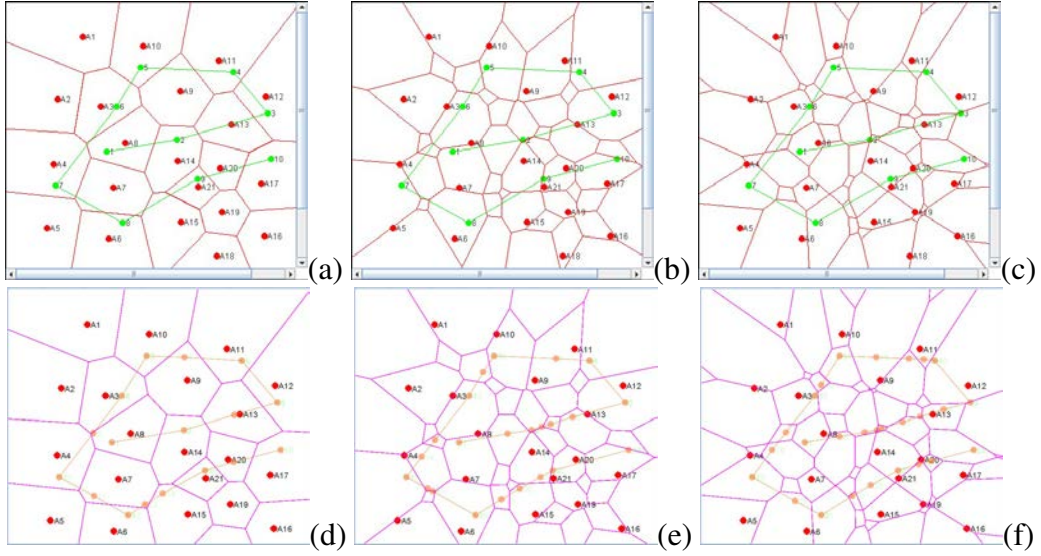


Fig. 4.4 $OkVD$:(a) O1VD; (b) O2VD; (c) O3VD with trajectory T_{16} and generators $P_M = \{p_{M.1}, \dots, p_{M.21}\}$ (labelled from A1 to A21); (d) topologically interpolated T_{16} with O1VD; (e) topologically interpolated T_{16} with O2VD; (f) topologically interpolated T_{16} with O3VD.

Figure 4.4(a), (b), and (c) display screen captures of visualisation tools with HOVD for a given POIs P_M and trajectory data T_{16} . Figure 4.4(a) shows O1VD, Figure 4.4(b) depicts O2VD and Figure 4.4(c) displays O3VD. Figure 4.4(d), (e) and (f) display screen captures of different topologically interpolated trajectory datasets based on $OkVD$. For topologically interpolated trajectory information, if trajectories path HOI regions that located empty data, then they are calculated for the trajectories data node into this empty region area and added into trajectories. This approach is balancing trajectories through different order k .

4.2.3 Discussion

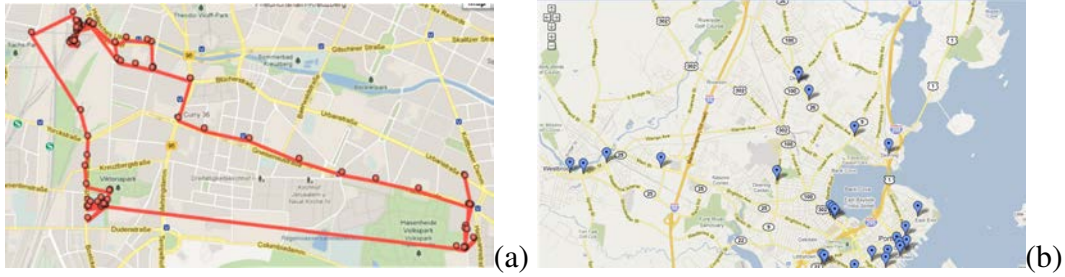


Fig. 4.5 Examples of real datasets: (a) GPS datasets; (b) picture targets from Facebook.

As geospatial data can be collected in different ways, it can be too dense as in Figure 4.5(a) and too coarse as in Figure 4.5(b). Depending on the trajectory data a user selects to analyse HOI, topologically simplified can filter the trajectory datasets when they are densely recorded. It restructures trajectory datasets based on topological information. Topologically interpolated can extend trajectory datasets when they are sparsely recorded, it reorganises trajectory datasets based on topological information. These approaches have been introduced to allow the user to choose a better representation to analyse moving objects with different dataset requirements.

In this research, we are collecting datasets from tour groups traveling around Cairns city for emergency management and tourism management. The datasets used in our research are irregular with different sampling rates. Some datasets are too densely recorded whilst others are too coarsely sampled. We introduce two input trajectory preprocessing approaches in Section 4.2 in order to handle this irregularity to achieve aim 2, and to better achieve our aims 3-7. For instance, the trajectory data is collected by GPS devices with different tour groups. The GPS data in Figure 4.5(a) is basically too dense. Thus, we are using the topologically simplified approaches as the main approach for data preprocessing to analyse trajectory HOI in this research. However, any of these two approaches could be used to preprocess to remove the irregular and unnecessary details.

Chapter 5

Visual Analytics Techniques

This chapter introduces five different VA techniques for Mul-HOI. They are higher-order parallel coordinate, higher-order spider, higher-order radar, higher-order treemaps, and higher-order DNA.

5.1 Higher-Order Parallel Coordinate (HOPC)

Original Parallel Coordinates: Original parallel coordinates (PCs) are widely adopted for visualising high-dimensional geometry and analysing multivariate (multiple targets) data. It shows a set of points in an n -dimensional space data in a 2-dimensional space as special node link diagrams, in which nodes are on the parallel axes and the edges are those polylines linking the nodes on two neighbouring axes. Since the introduction of PCs charts [99], many enhancements of PCs have been developed [100–104]. This approach is a perfect candidate to display Mul-HOI and relationships between POIs and trajectories of k NN and k OR information.

Proposed HOPC: PCs are useful for displaying a trend over different values. The proposed HOPC offers a very similar to original method in terms of the analysis of data structure. Proposed data structure displays HOI with different k^{th} order and the changing trend of geometrical information and topological information of relationship with POIs and directional information of the position of the POIs. Quantitative and qualitative

geometrical dissimilarity measurements are proposed with regard to k NN information of Mul-HOI. The proposed HOPC addresses research Aims 3, 4, and 5 by providing VA tools for quantitative/qualitative geometrical/topological/directional HOI with multivariate datasets for moving objects. In addition, a geometrical dissimilarity measurement addresses research Aim 6.

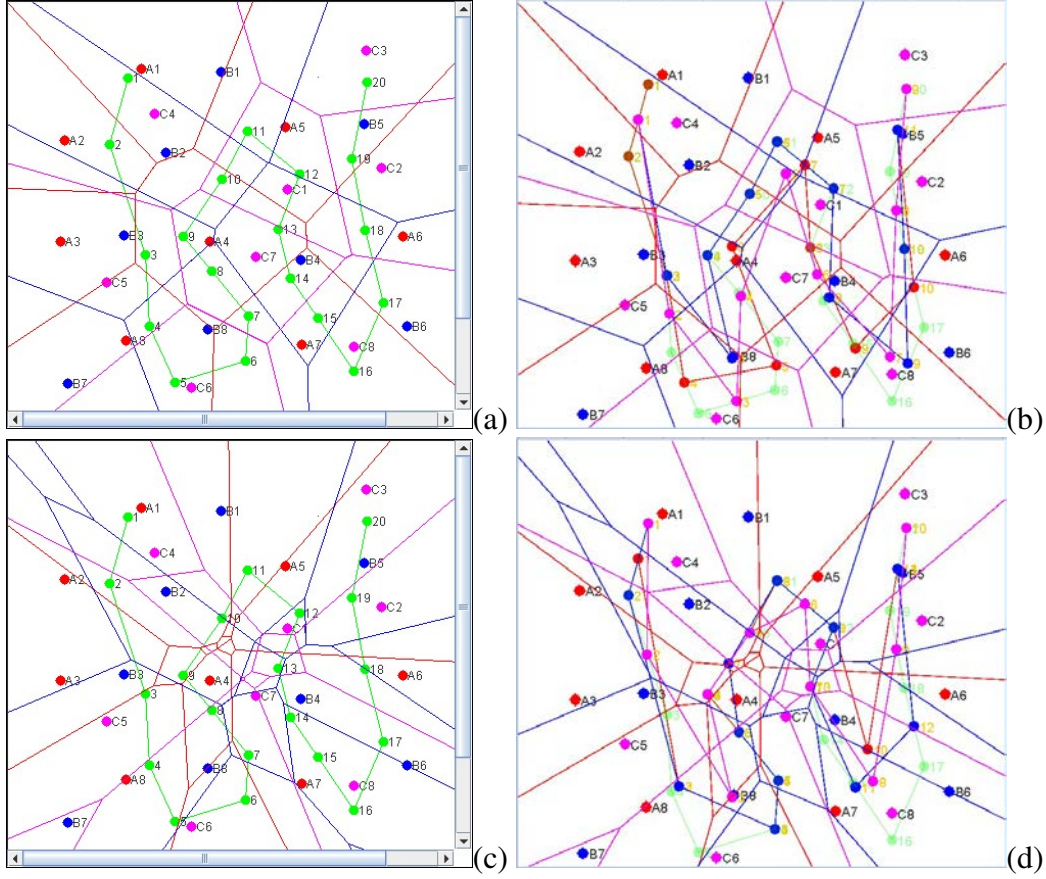


Fig. 5.1 OkVDs: (a) O1VD and (c) O5VD with generators $P_N = \{p_{N.1}, \dots, p_{N.8}\}$ in red (labelled from A1 to A8), $P_O = \{p_{O.1}, \dots, p_{O.8}\}$ in blue (labelled from B1 to B8), $P_P = \{p_{P.1}, \dots, p_{P.8}\}$ in magenta (labelled from C1 to C8) and trajectory data $T_{17} = \{t_{17.1}, \dots, t_{17.20}\}$ in green; (b) O1VD and (d) O5VD with generators, P_N for topologically simplified trajectory T_{17} in red, P_O for topologically simplified trajectory T_{17} in blue and P_P for topologically simplified trajectory T_{17} in magenta.

Figure 5.1 displays screenshots of the proposed VA suite with three generators P_N , P_O , and P_P , and a trajectory dataset T_{17} . The Mul-HOI implementation enables users to change various k values to obtain different types of Mul-HOI from the trajectory data. Since different

5.1 Higher-Order Parallel Coordinate (HOPC)

OkVDs are drawn and visualised from the same set of datasets P_N , P_O , and P_P , users are able to interact with the program suite to retrieve different data points for any user-interested location within the set of query points [3] on the fly. P_N and its corresponding Voronoi diagrams are represented in red, P_O and its corresponding Voronoi diagrams are represented in blue, and P_P and its corresponding Voronoi diagrams are shown in magenta. These Voronoi diagrams are drawn based on the geometrical HOI. Geometrical HOI is computed based on the Euclidean distance of the trajectory data points to a set of nearest generators. O1VD and O5VD of the trajectory data are shown in Figure 5.1(a) and 5.1(c) in green. Figure 5.1(b) and 5.1(d) display topologically simplified trajectories with multivariate generators' HOI. For the topologically simplified trajectories, the program will try every single data point (node) in the trajectory, cluster the data points possessing the same topological HOI (falling into the same Voronoi region), and represent them as a simplified centre node. A topologically simplified trajectory for P_N is presented in red, and a topologically simplified trajectory for P_O is presented in blue, and a topologically simplified trajectory for P_P is presented in magenta.

5.1.1 Geometrical PC

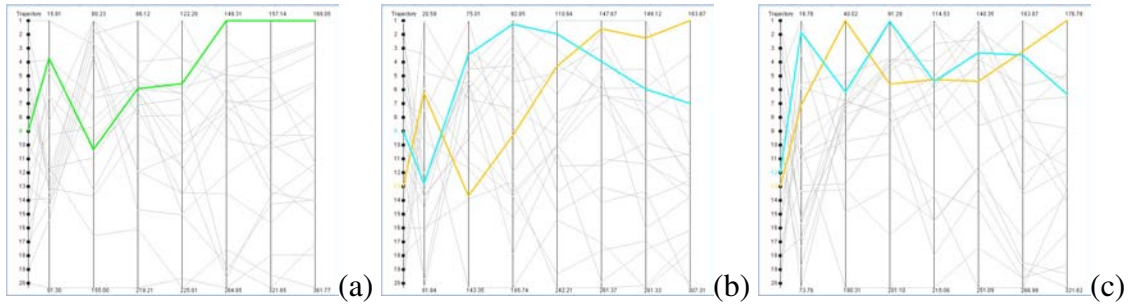


Fig. 5.2 Geometrical PCMMs (x-axis: minimum (top) and maximum (bottom) of the distance to Order- k generators; y-axis: trajectory node ID): (a) Geometrical PCs - P_N ; (b) Geometrical PCs - P_O ; (c) Geometrical PCs - P_P based on the trajectory data T_{17} . Quantitative PCMM is shown in orange and qualitative PCMM is shown in cyan. When quantitative and qualitative PCMMs are the same, it is shown in green.

Parallel Coordinates Multivariate Measure (PCMM) for mining trajectory HOI is defined based on the HOI information discussed above. Figure 5.2 exhibits geometrical PCs with distance and relationship between trajectory and different generators. For each data point of the trajectory (T_{17}) and three POIs (P_N, P_O, P_P), we calculate the distance between each single POI and the trajectory with different order k (x-axis of the diagram) to draw the geometrical PCs. Each column (vertical bar) from left to right in the figure shows Order-1 to Order- k HOI. For each data point in the trajectory, the distance to the k NN is calculated using v_{Short} as shown in Equation 5.1 to determine a qualitative PCMM. For each data point in the HOI trajectory of the three POIs, the distances of the data point to the 1NN to k NN HOI are used to calculate v_{Short} . If the data point has the shortest distance from HOI in P_N , this trajectory point is calculated and it determines a quantitative PCMM for P_N . Quantitative PCMM is shown in orange, and qualitative PCMM is shown in cyan. For the same quantitative and qualitative PCMM, it is shown in green. In Equation 5.1, Min and Max represent the minimum and maximum distances between any point in the trajectory to the k NN POIs. The $newMax$ and $newMin$ are scaling factors which can be chosen by users to scale the PCMM into the range of values as desired. $newMin$ is set to 0 and $newMax$ is set to 1 in this thesis, hence v_{Short} will be scaled between 0 and 1.

$$v_{Short} = \frac{v - Min}{Max - Min} (newMax - newMin) + newMin. \quad (5.1)$$

5.1.2 Topological PC

When considering the topological HOI of the trajectory, Figure 5.3 shows PCMMs of the topological information exhibiting relationships between trajectory and generators (POIs). For each data point of the HOI trajectory (T_{17}) with respect to the three sets of POIs (P_N, P_O, P_P), we calculate the relationship between each trajectory data point and single POI with different order k (x-axis of the diagram) to draw topological PCs. Each column from left to right shows the Order-1 to Order- k information, respectively. Each trajectory data goes through each column by calculating the relationship of each trajectory to POIs, then each trajectory data draws the relationship to each k column.

5.1 Higher-Order Parallel Coordinate (HOPC)

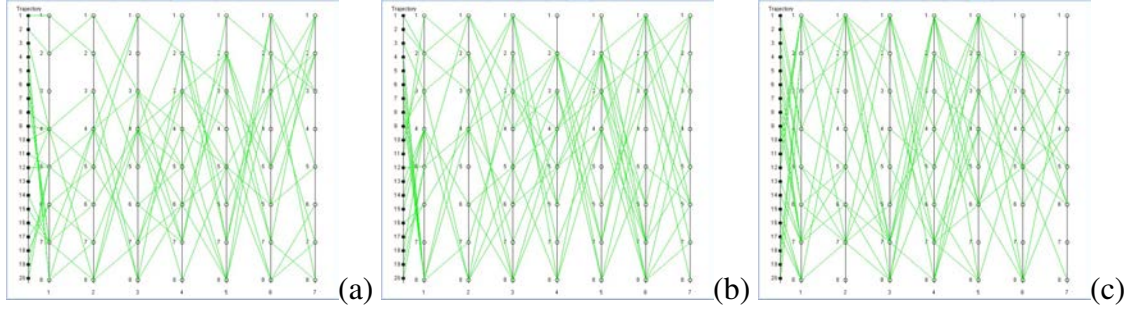


Fig. 5.3 Topological PCMMs (x-axis: Order- k ; y-axis: trajectory node ID): (a) Topological PCs - P_N ; (b) Topological PCs - P_O ; (c) Topological PCs - P_P based on the trajectory data T_{17} .

5.1.3 Directional PC

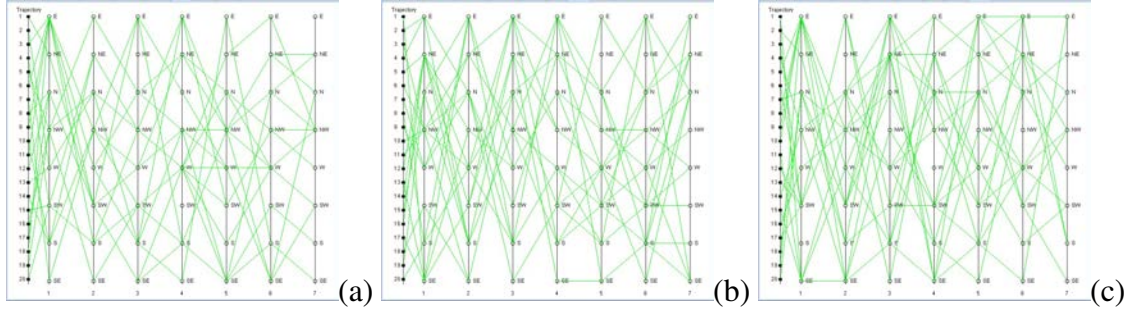


Fig. 5.4 Directional PCMMs (x-axis: Order- k ; y-axis: trajectory node ID): (a) Directional PCs - P_N ; (b) Directional PCs - P_O ; (c) Directional PCs - P_P based on the trajectory data T_{17} .

Considering the directional HOI trajectory, PCMMs of directional information for mining trajectory exhibit directions (East (E), West (W), North (N), South (S), Northeast (NE), Southeast (SE), Northwest (NW), Southwest (SW)) between trajectory and generators as shown in Figure 5.4. For each single data point of the HOI trajectory (T_{17}) of three POIs (P_N , P_O , P_P), we calculate directional information between each single POI and trajectory with different order k (x-axis of the diagram) to draw directional PCs. Each column from left to right shows the Order-1 to Order- k information, respectively. Each trajectory data goes through each column by calculating directional information of each trajectory to POIs, then each trajectory data draws a direction to each k column. In conclusion, we are able to reveal three different categories of HOI with PCMMs. The running time of this approach is linear to

the number of trajectory nodes. This approach is much faster than other existing approaches and is able to provide more details and richer information compared to other VA approaches.

5.1.4 Summary and Discussion

This section proposes a unified data structure for multivariate datasets that can support Mul-HOI; a VA approach for topologically simplified HOI of trajectory datasets; and PCMM for quantified Mul-HOI with regard to k NN and k OR information. Those three approaches enable users to analyse large volumes of trajectories with different POIs, related to the trajectories to understand the relationship of POIs and trajectory HOI. We test our proposed approaches with a case study explaining the usefulness and applicability of our approaches in Chapter 7.

Table 5.1 Research contribution for HOPC.

	Univariate/ Multivariate	Trajectory Support	Interactivity	Quantitative/ Qualitative	Higher-Order Information (HOI)			Higher-Order Dissimilarity Measurement			Higher-Order Data Mining
					Geo	Top	Dir	Geo	Top	Dir	
HOPC	Uni/Mul	Yes	Yes	Quan/Qual	Yes	Yes	Yes	Yes	No	No	No

Table 5.1 shows the contribution of methodology HOPC that addresses the research gap. HOPC can fill the research gap of geometrical, topological, and directional HOI and can provide a geometrical dissimilarity measurement.

5.2 Higher-Order Spider (HOS)

Original Spider Chart: Original spider chart is a graphical method of displaying multivariate data in the form of a two-dimensional chart of three or more quantitative variables represented on axes starting from the same point. It is a good candidate for providing directional information visualisation for comparison of multivariate datasets. Spider charts have been used in many different areas to measure customer satisfaction metrics [105] and to evaluate mobile service [106].

Proposed HOS: Proposed spider chart is a new graphical method of displaying trajectory HOI in the form of a two-dimensional chart of eight different directions on axes starting from the same point based on the Min-Max normalisation for k^{th} HOI. Proposed HOS can answer research Aims 4 and 5 by providing VA tools for quantitative/qualitative topological and directional HOI with multivariate datasets for moving objects. Also, two proposed metrics based on the Min-Max normalisation for k^{th} HOI can achieve part of research Aim 6.

5.2.1 Directional Spider Chart (DSC)

In this section, we present a visualisation approach for displaying directional HOI with 8 different directions: $dir = \text{East, West, South, North, Southeast, Southwest, Northeast, Northwest}$. The Min-Max normalisation is one of the most widely used normalisations that converts quantitative values into a certain range $[newMax, newMin]$. In order to visualise relative ratios of directional information, the Min-Max normalisation (Equation 5.1) has been used as default here.

A Directional Spider Chart (DSC) is based on Min-Max normalised values of trajectory HOI. First of all, directional information for each trajectory node $t_k \in (t \text{ or } t')$ (the original trajectory node t for directional-geometrical HOI DSC whilst t' for topologically simplified directional-topological HOI DSC) is calculated and accumulated for that direction $d \in dir \neq 1$. Once cardinality of each direction is calculated, then it is normalised into $[0,1]$ for relative visualisation.

Figure 5.5(a) and 5.5(b) display screenshots of two trajectory datasets T_{18} and T_{19} and generators P_Q . T_{18} is represented in green and T_{19} is represented in blue. P_Q is represented in red. O1VD and O3VD are shown. Geometrical HOI is computed based on the Euclidean distance of the trajectory data points to a set of nearest generators. Directional HOI is about relative directional positioning of trajectory nodes to their kNN generators P_Q . In many cases, the number ($t = \{t_1, \dots, t_n\}$) of trajectory nodes is large and many neighbouring trajectory nodes fall in the same kOR . A trajectory $t = \{t_1, \dots, t_n\}$ can be topological simplified into $t' = \{t_p, t_{p+1}, \dots, t_q\}$ where $t_p \in t$ for $p = 1, \dots, q$ and $kOR(t_p) \neq kOR(t_{p+1})$. Figure 5.5(c)

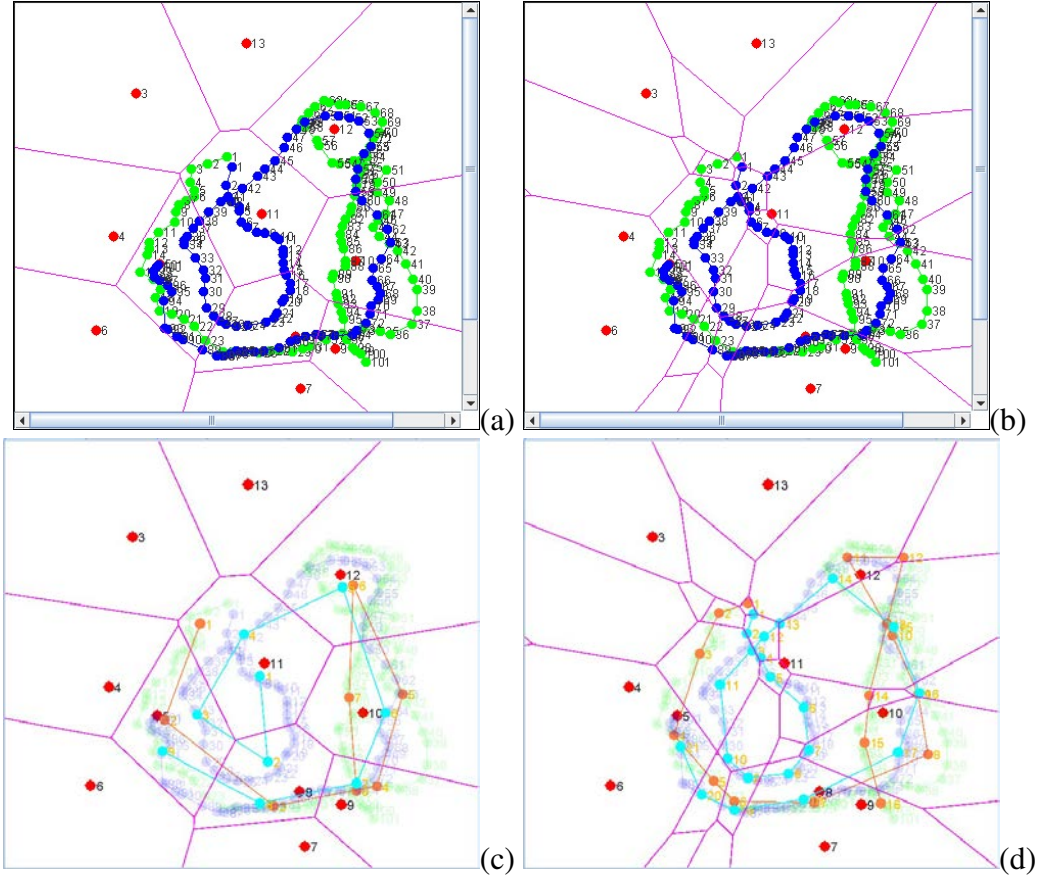


Fig. 5.5 OkVDs with two trajectories $T_{18} = \{t_{18.1}, \dots, t_{18.101}\}$, $T_{19} = \{t_{19.1}, \dots, t_{19.101}\}$ and generators $P_Q = \{p_{Q.1}, \dots, p_{Q.13}\}$ (labelled from A1 to A13): (a) O1VD; (b) O3VD; (c) topologically simplified T_{18} and T_{19} with O1VD; (d) O3VD with two topologically simplified trajectory datasets T_{18} and T_{19} .

and 5.5(d) display topologically simplified trajectories in orange for T_{18} and light blue for T_{19} .

Figure 5.6(a)-5.6(d) show two trajectories T_{18} and T_{19} . T_{18} is shown in green while T_{19} is in blue. The area of colour covered in DSC provides an easy and VA way to compare the directional HOI. For instance, T_{19} exhibits west intensive directional information for lower order until 5th order, but as the order grows to 10 (shown in Figure 5.6(d)) the east direction becomes the most dominant. North consistently shows dominance for T_{18} . Also, the more the two areas overlap each other, the more similar the directional HOI of the two trajectories. That is, the two trajectories T_{18} and T_{19} are most dissimilar in 1st order information (shown in

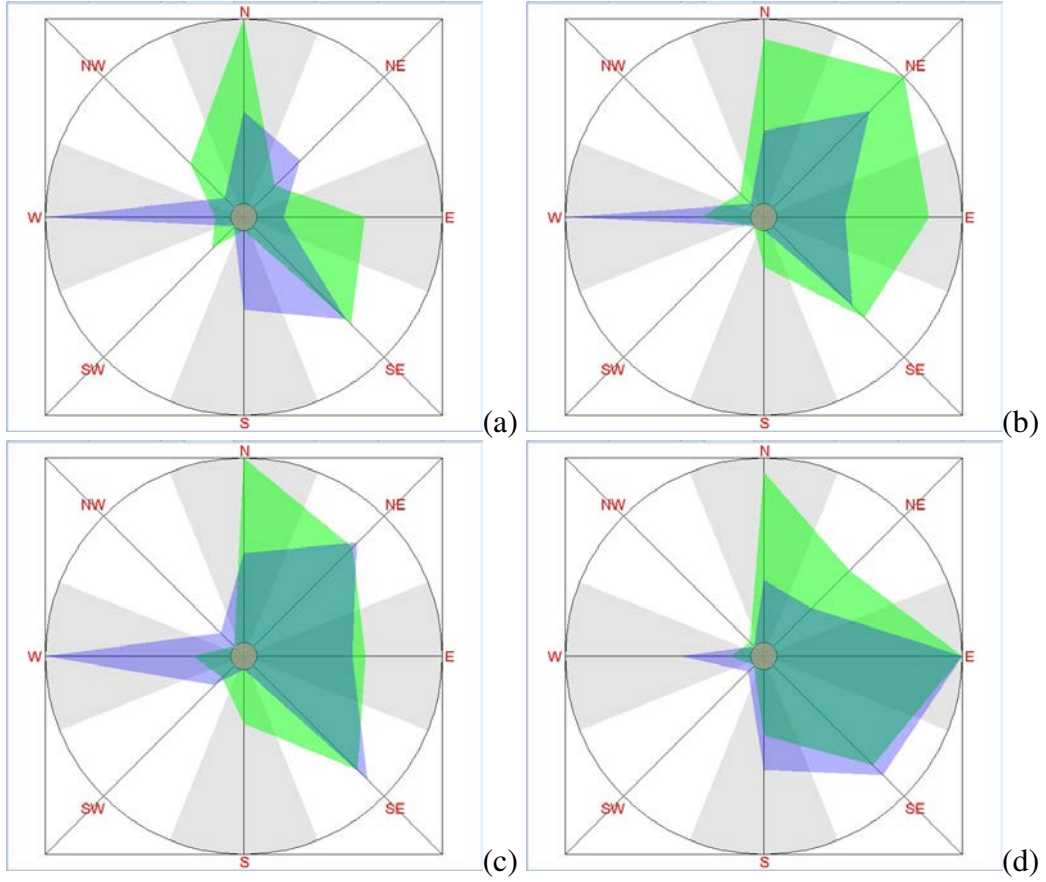


Fig. 5.6 Directional-geometrical DSCs: (a) O1VD DSC; (b) O3VD DSC; (c) O5VD DSC; (d) O10VD DSC with two trajectory datasets T_{18} , T_{19} and generators P_Q .

Figure 5.6(a)), while they are most similar in 10^{th} order information (shown in Figure 5.6(d)). However, this trend is not evident in directional-topological DSCs shown in Figure 5.7.

5.2.2 Directional Balance Ratio (DBR)

In order to measure a quantitative comparison ratio for directional HOI of trajectories, we introduce Directional Balance Ratio (DBR). DBR is the average of Directional Ratios (DRs) where a DR is for a single pair of opposite directions, e.g., east and west. Given a trajectory $T = \{t_1, \dots, t_n\}$ of n nodes and a set $P = \{p_1, \dots, p_m\}$ of m generators, let $kNN(t_i)$ ($t_i \in T$) return a subset $P_s \subset P$ of k^{th} order nearest neighbours, and let each directional function $kEast(T, P)$ (the rest of directional functions, $kWest(T, P)$, $kNorth(T, P)$, $kSouth(T, P)$,

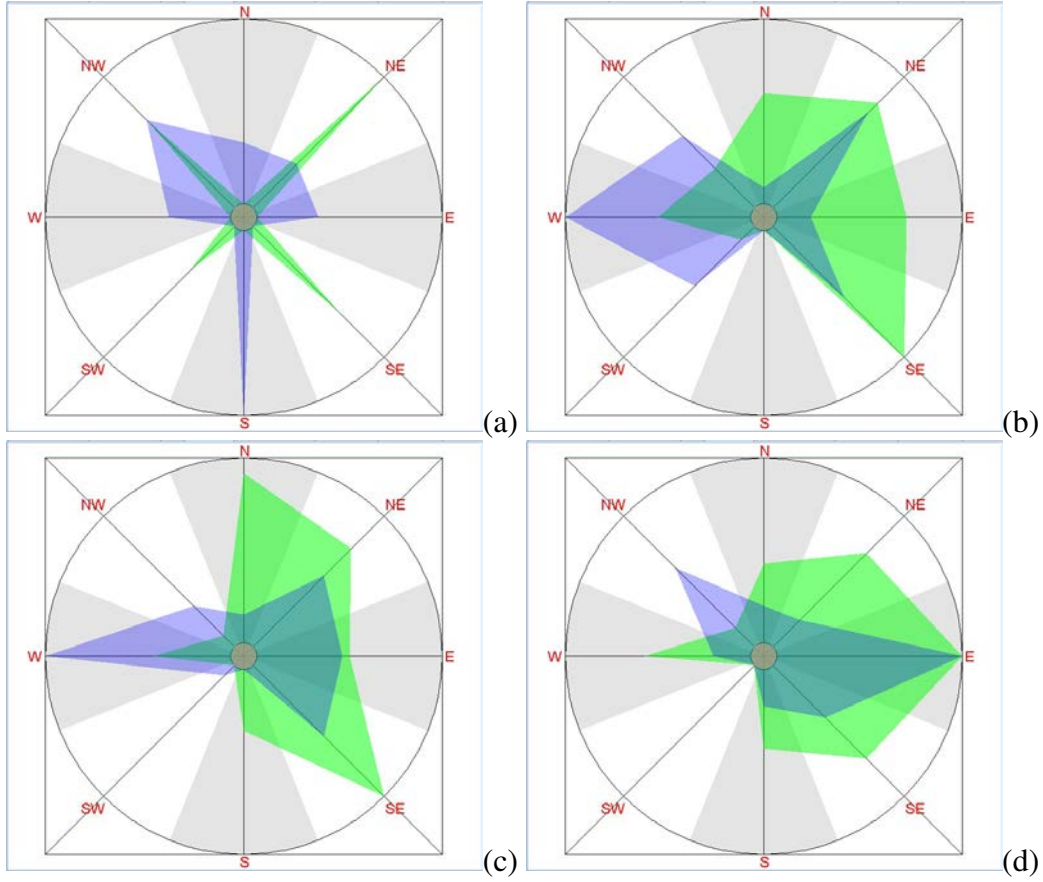


Fig. 5.7 Directional-topological DSCs: (a) O1VD DSC; (b) O3VD DSC; (c) O5VD DSC; (d) O10VD DSC with two topologically simplified trajectory datasets T_{18} , T_{19} and generators P_Q .

$kNW(T, P)$, $kSE(T, P)$, $kNE(T, P)$, $kSW(T, P)$ could be similarly defined) return the number of $t_i \in T$ that lies a western side of its k^{th} order nearest neighbours $\in P$.

For a given k^{th} order, a trajectory T and a generator P , k^{th} order DRs are defined as follows:

$$kDR_{ew} = \begin{cases} 0 & \text{if } \max() \text{ is } 0, \\ \frac{\min(kEast(T, P), kWest(T, P))}{\max(kEast(T, P), kWest(T, P))} & \text{if } \max() \text{ is not } 0, \end{cases} \quad (5.2)$$

$$kDR_{ns} = \begin{cases} 0 & \text{if } \max() \text{ is } 0, \\ \frac{\min(kNorth(T, P), kSouth(T, P))}{\max(kNorth(T, P), kSouth(T, P))} & \text{if } \max() \text{ is not } 0, \end{cases} \quad (5.3)$$

$$kDR_{nwse} = \begin{cases} 0 & \text{if } \max() \text{ is } 0, \\ \frac{\min(kNW(T,P), kSE(T,P))}{\max(kNW(T,P), kSE(T,P))} & \text{if } \max() \text{ is not } 0, \end{cases} \quad (5.4)$$

$$kDR_{nesw} = \begin{cases} 0 & \text{if } \max() \text{ is } 0, \\ \frac{\min(kNE(T,P), kSW(T,P))}{\max(kNE(T,P), kSW(T,P))} & \text{if } \max() \text{ is not } 0, \end{cases} \quad (5.5)$$

where, $\min(a, b)$ returns the minimum of a and b whilst $\max(a, b)$ returns the maximum of a and b . k^{th} order DBR is defined as below:

$$kDBR = \frac{(kDR_{ew} + kDR_{ns} + kDR_{nwse} + kDR_{nesw})}{4} \quad (5.6)$$

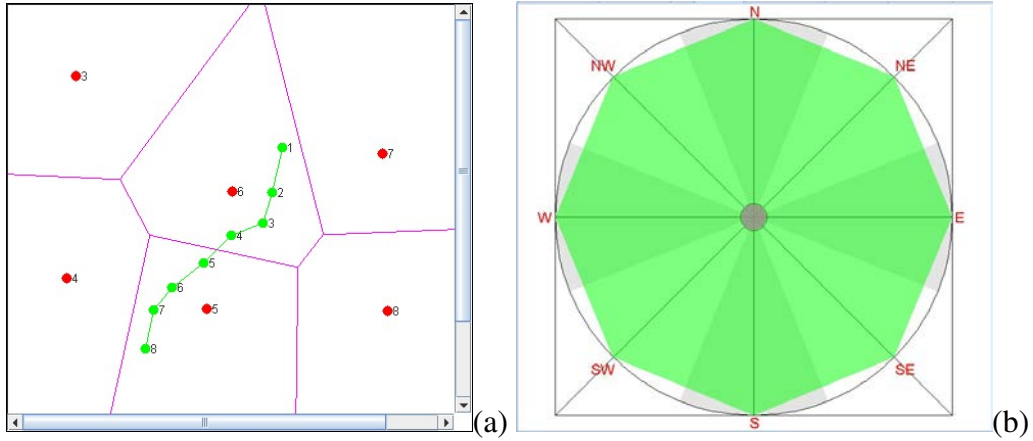


Fig. 5.8 A well-balanced example: (a) O1VD; (b) DSC of O1VD based on directional-geometrical HOI DBR of trajectory datasets $T_{20} = \{t_{20.1}, \dots, t_{20.8}\}$ and generators $P_R = \{p_{R.1}, \dots, p_{R.6}\}$.

DBR values vary between 0 and 1, where lower values imply directionally skewed distributions while higher values imply directionally well balanced distributions. Note that a DBR value 1 means the most balanced distribution as shown in Figure 5.8.

Figure 5.9 shows DBRs of the two trajectory datasets T_{18} and T_{19} to illustrate the distribution of DBR values with regard to the higher-order k . As you can see from Figure 5.9(a), the trajectory T_{19} exhibits a much more balanced distribution than T_{18} for both directional-geometrical HOI and directional-topological HOI. As k grows, DBRs for T_{19} tend to increase, and when $k = 7$, T_{19} exhibits the highest directional balance. However, T_{18} shows the most

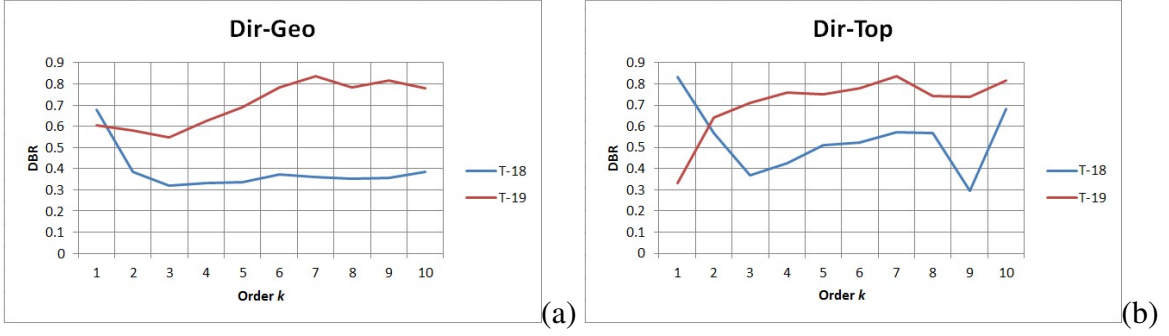


Fig. 5.9 Line graphs: (a) DBRs for directional-geometrical HOI; (b) DBRs for directional-topological HOI of trajectory datasets T_{18} and T_{19} .

balanced distribution when $k = 1$, and DBR values for T_{18} substantially drop when $k = 2$ and stay constant for the rest of k values except there is a sudden drop in $k = 9$ for directional-topological HOI. These line graphs evidently demonstrate that T_{19} shows a more balanced directional distribution than T_{18} .

5.2.3 Directional Dissimilarity Measure (DDM)

Based on directional HOI, we can define Directional Dissimilarity Measure (DDM) for mining trajectories exhibiting similar directional HOI. For a given order k and generators P , k^{th} order DDM of two trajectories T_{18} and T_{19} is defined as below:

$$kDDM(T_{18}, T_{19}) = \sqrt{kDDMS(T_{18}, T_{19})}, \quad (5.7)$$

where $kDDMS(T_{18}, T_{19})$ is defined as below:

$$\begin{aligned}
kDDMS(T_{18}, T_{19}) = & (kEast(T_{18}, P) - kEast(T_{19}, P))^2 \\
& + (kWest(T_{18}, P) - kWest(T_{19}, P))^2 \\
& + (kNorth(T_{18}, P) - kNorth(T_{19}, P))^2 \\
& + (kSouth(T_{18}, P) - kSouth(T_{19}, P))^2 \\
& + (kNW(T_{18}, P) - kNW(T_{19}, P))^2 \\
& + (kSE(T_{18}, P) - kSE(T_{19}, P))^2 \\
& + (kNE(T_{18}, P) - kNE(T_{19}, P))^2 \\
& + (kSW(T_{18}, P) - kSW(T_{19}, P))^2.
\end{aligned} \tag{5.8}$$

Smaller values of $kDDM(T_{18}, T_{19})$ mean that T_{18} and T_{19} are more similar while larger values imply less similarity.

5.2.4 Summary and Discussion

Due to the complexity of trajectory dataset and related data analysis, there is a need to develop new and innovative ways to allow researchers and data analysts to reveal hidden information and knowledge embedded in trajectory datasets. While most of the existing research efforts focus on geometrical HOI and topological HOI, studies related to directional HOI of trajectory data are limited. This section proposes a new VA approach to visualise directional HOI. In particular, by combining directional HOI with geometrical HOI and combining directional HOI with topological HOI, we create a DSC. This chart provides a means to visually evaluate the balance of a trajectory dataset with reference to different directions as well as comparing the directional HOI and balance among trajectory datasets. DBR and DDM are also introduced in this approach providing a quantitative way to compare the balance of directional HOI of trajectories and contrast the similarity for multivariate trajectory datasets. The main result of this approach is, on one hand, both of these are pioneering ways to investigate directional information with trajectory HOI, on the other hand,

the established links about where (space), when (time), what (objects), and which (direction) analysis for HOI.

Table 5.2 Research contribution for HOS.

	Univariate/ Multivariate	Trajectory Support	Interactivity	Quantitative/ Qualitative	Higher-Order Information (HOI)			Higher-Order Dissimilarity Measurement			Higher-Order Data Mining
					Geo	Top	Dir	Geo	Top	Dir	
HOS	Uni/Mul	Yes	Yes	Qual	No	No	Yes	No	No	Yes	No

Table 5.2 shows the contribution of methodology HOS that can address the research gap. HOS method can fill the research gap of directional HOI and can provide a directional dissimilarity measurement.

5.3 Higher-Order Radar (HOR)

Original Radar Chart: Original radar chart emerges as a promising candidate for displaying multivariate data in the form of a two-dimensional chart of three or more quantitative variables represented on axes starting from the same point. It is useful for multivariate information visualisation. There is a strong need to represent all three types of HOI in one visualisation and the radar chart is a good candidate to represent all these since it has been used in food material and food products [107], health care [108], network [109] and multimedia services [110, 111] in the literature.

Proposed HOR: The proposed radar chart is used to display k NN Mul-HOI on a radar chart for geometrical, topological, and directional information. Clusters of HOI from HOI mining methodology can also be displayed in this HOR. The proposed HOR can answer research Aims 3, 4 and 5 with VA tools for all three types of HOIs with multivariate datasets for moving objects. Higher-Order k -Means from the HOI mining methodology can locate top- k for k NN and k OR information, which can achieve research Aim 7.

5.3.1 Higher-Order Radar Chart (HORC)

In this section, we present a VA approach called HORC. To create a HORC, first of all, geometrical information for each trajectory node $t_k \in T$ (original trajectory T for geometrical HORC) is calculated for that distance to POIs. Once the cardinality of each distance is calculated, it is normalised into $[0,1]$ for relative visualisation based on Equation 5.1. Then, directional information for each trajectory node $d_k \in D$ (original trajectory D for directional HORC) is determined for that direction from POIs.

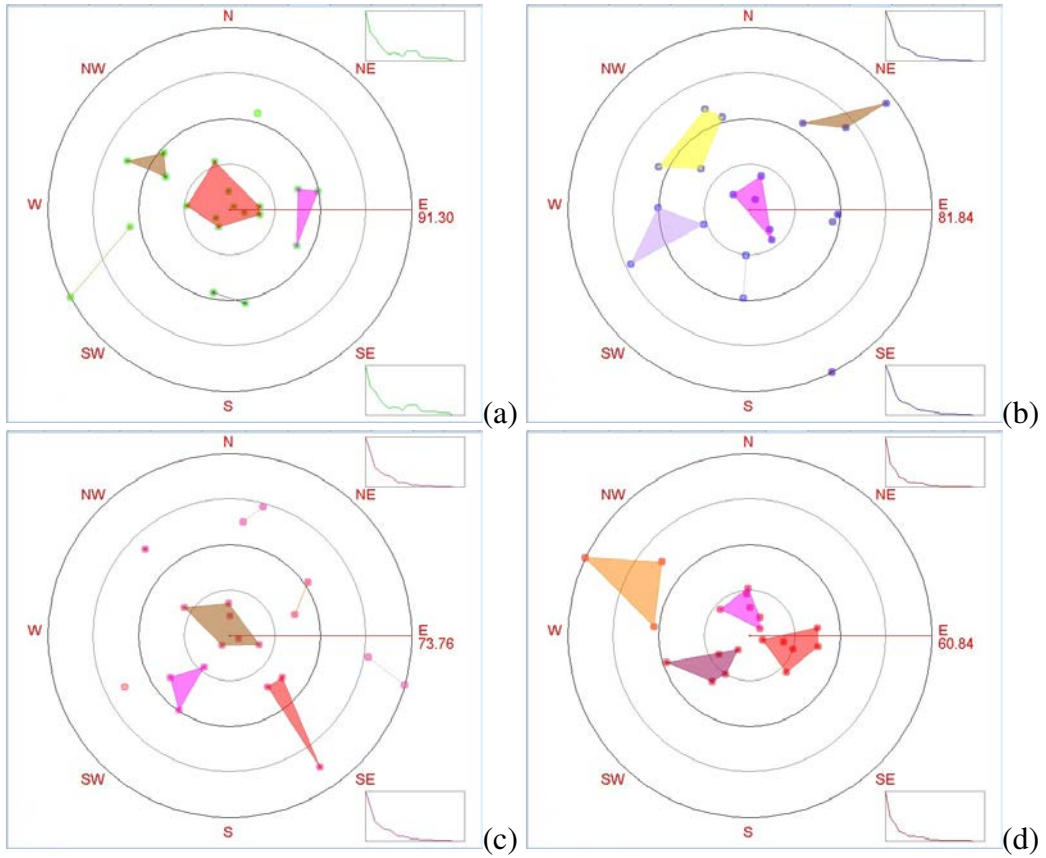


Fig. 5.10 HORCs with the trajectory T_{17} used in Figure 5.1: (a) O1VD HORC based on P_N ; (b) O1VD HORC based on P_O ; (c) O1VD HORC based on P_P ; (d) O1VD HORC based on all P_N , P_O , and P_P together.

Figure 5.10 and Figure 5.11 show HORCs of the four different groups of datasets used in Figure 5.1. The radar chart consists of four circles of different sizes sharing the same centre. The ratio of the radii of the circles is 1:2:3:4. We use this to measure

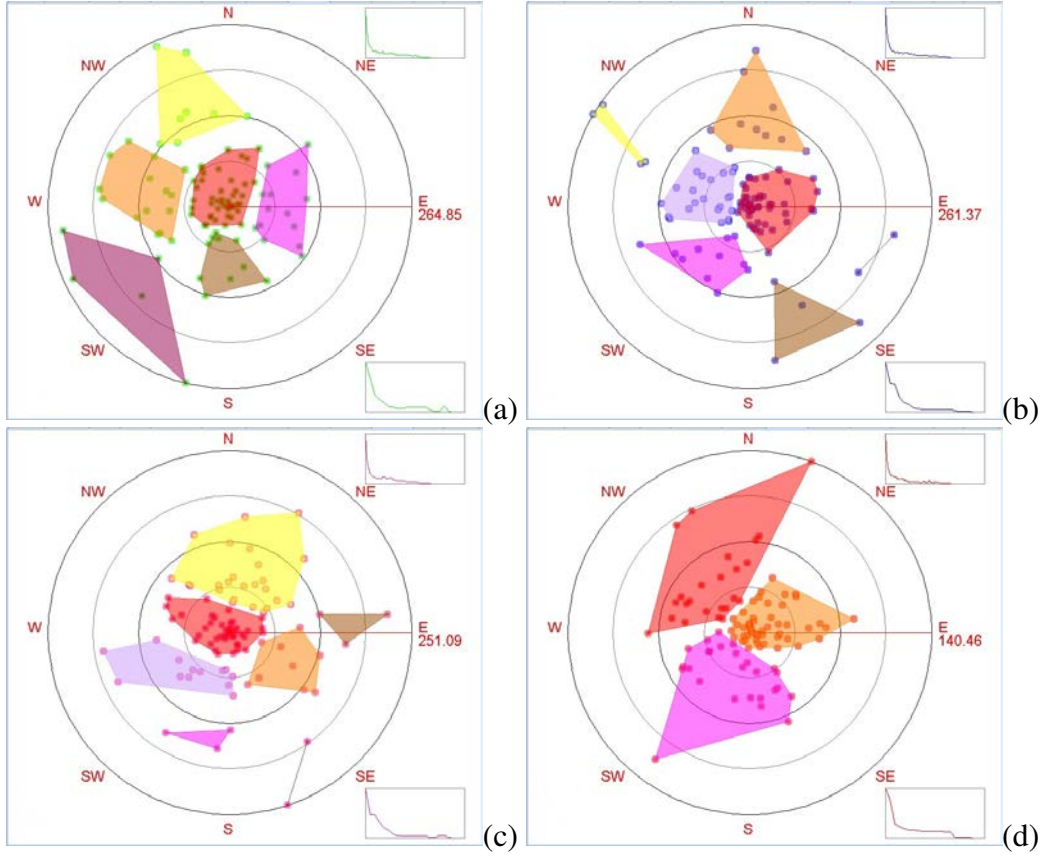


Fig. 5.11 HORCs with the trajectory T_{17} used in Figure 5.1: (a) O5VD HORC based on P_N ; (b) O5VD HORC based on P_O ; (c) O5VD HORC based on P_P ; (d) O5VD HORC based on all P_N , P_O , and P_P together.

the Min-Max normalisation and to represent normalised values into this radar chart. At the *East* side of circle, we display the *Max* distance of the Min-Max normalisation to measure and compare with trajectories. The k NN POIs of each trajectory node are shown inside of these circles with different colours to represent POIs of distance (geometrical), clusters (topological relationship), and direction (directional) information. Therefore, HORCs represent geometrical, topological, and directional HOI information at the same time.

5.3.2 Higher-Order k -Means (HO k -Means)

In order to cluster quantitative top- k HOI for higher-order trajectory data mining, Higher-Order k -Means (HO k -Means) is introduced. HO k -Means is a k -Means approach to calculate

an average distance up to k^{th} order HOI. First, each trajectory node finds k^{th} order POIs; then, HOk -Means calculates all k^{th} order POIs “*Mean*” to find the mean of top- k of this trajectory HOI. The higher-order cluster centre T' is defined as below:

$$T' = \sum_{i=1}^n Means(kNN(t_i)). \quad (5.9)$$

For example, given a trajectory $T = \{t_1, \dots, t_n\}$ of n nodes and a set $P = \{p_1, \dots, p_m\}$ of m generators, let $kNN(t_i)$ ($t_i \in T$) return a subset $P_i \subset P$ of k^{th} order nearest neighbours, and find the “*mean*” of P_i , and let each trajectory node $t_i \in T$ find new T' which includes centroids of all clusters. In the end, the Sum Squared Error (SSE) for all values of cluster number k is computed. The SSE is defined as the sum of the squared distance between each member of the cluster and its centroid. Mathematically it is defined as:

$$SSE = \sum_{i=1}^k \sum_{j=1}^n \left\| t_i^{(j)} - T'_j \right\|^2, \quad (5.10)$$

where $\left\| t_i^{(j)} - T'_j \right\|^2$ is a chosen distance measure between a trajectory data node $t_i^{(j)}$ and the higher-order cluster centre T'_j , which is an indicator of the distance of the n data points from their corresponding higher-order cluster centres. This method not only minimises a cluster’s centroid but also considers HOI of each trajectory. On the corners of HORCs, the line charts for SSE are for different k values. At the top right corner is a line chart using the original elbow method [112], on the bottom right is a line chart using the HOk -Means method with trajectory considered.

Figure 5.10(a) shows HORC based on the O1VD for P_N , and HOI of T_{17} is displayed in green. Figure 5.10(a) shows that, based on O1VD, most of the POIs are very close to trajectory T_{17} , almost half of POIs distance are inside the innermost circle, that is under 25% of 91.30 km. Generally, most POIs sit at the west side of trajectory T_{17} . Figure 5.10(b) displays O1VD HORC for P_O , and HOI of T_{17} is displayed in blue. Figure 5.10(b) shows that, based on O1VD, most of the POIs are half distance (about 50%) of maximum distance range which is 81.84km. Interestingly, most of POIs are sitting at the west side of trajectory T_{17} similar to P_N . Figure 5.10(c) presents O1VD HORC for P_P . HOI of T_{17} is displayed in

magenta. Figure 5.10(c) shows that based on O1VD, most of the POIs are under middle circle (about 50%) of maximum distance range which is 76.76km. Obviously, most of POIs are sitting at the east side of trajectory T_{17} . Figure 5.10(d) shows O1VD HORC for all P_N , P_O , and P_P together, and HOI of T_{17} is displayed in red. Based on O1VD, most of the POIs are located inside the middle circle that is under 50% of 60.84km, and most of POIs are sitting at the west side of trajectory T_{17} .

Figure 5.11 displays O5VD HORCs which consider the five nearest POIs for each node. Using the proposed HOK-Means clustering method, considering each trajectory separately, 6 clusters are identified for P_N as in Figure 5.11(a), 7 clusters are identified for P_O as in Figure 5.11(b), and 7 clusters are identified for P_P as in Figure 5.11(c) based on the HOI and O5VD. Again, when considering the O5VD for all generators together, 3 clusters are identified for P_N , P_O , and P_P as in Figure 5.11(d).

5.3.3 Summary and Discussion

This section proposes a new VA approach to visualise HOI with regard to target generators and a new trajectory clustering method with trajectory HOI. In particular, by combining geometrical HOI with directional HOI and determining the optimal k , we propose a HORC and HOK-Means clustering. The chart is a new VA tool to explore both the geometrical and directional HOI at the same time. The HOK-Means clustering helps to discover hidden groups in the dataset with considering HOI. Such charts and clustering are very useful for complex situations when both geometrical HOI and directional HOI have to be considered at the same time. An example for this situation could be during natural disasters such as flooding and hurricane. Not only knowing the k -nearest emergency centres is important, but the direction of the centres should also be considered as the water or wind may move from one particular direction to another. HORC and the HOK-Means clustering provide information to allow management to allocate different levels of resources to different clusters. Experimental results demonstrate promising outcomes in Chapter 7 that support the usefulness and applicability of the proposed methods. Future work includes further experiments with large trajectory datasets and extensions to multivariate HOI.

5.4 Higher-Order Treemaps (HOTrM)

Table 5.3 Research contribution for HOR.

	Univariate/ Multivariate	Trajectory Support	Interactivity	Quantitative/ Qualitative	Higher-Order Information (HOI)			Higher-Order Dissimilarity Measurement			Higher-Order Data Mining
					Geo	Top	Dir	Geo	Top	Dir	
HOR	Uni/Mul	Yes	Yes	Quan/Qual	Yes	Yes	Yes	No	No	No	Yes

Table 5.3 shows the contribution of methodology HOR that addresses the research gap. The HOR method fills the research gaps of geometrical, topological and directional HOI and supports HOI data mining methodology.

5.4 Higher-Order Treemaps (HOTrM)

Original Treemaps: Treemaps possess a powerful layout, data processing, and rendering engine that offers a vast choice of customisation possibilities. Treemaps were originally designed to display hierarchical (tree-structured) data and visualise files on a computer hard drive to overcome the limitations of node and link diagrams [33]. A treemap works by dividing the display area into a nested sequence of rectangles whose areas correspond to an attribute of the dataset. A treemap is constructed by recursive subdivision of a rectangle to some sub-rectangles corresponding to the size of the node. A key component of a treemap is the algorithm used to create the nested rectangles that make up the map. Colour and annotations can be used with treemaps to show additional information. Several alternative layout algorithms [33] have recently been proposed to facilitate different applications, in particular, by controlling the aspect ratios of the rectangles that compose the treemap.

Proposed HOTrM: The proposed treemap displays 7 different types of treemap methods to support k NN and k OR information. It offers a full set of the most successful tiling algorithms with different aspect ratios of the composing shapes while balancing their ability to maintain order and their stability with respect to changes in the data. We compare and contrast the 7 different treemap methods and evaluate them to select the

best treemap structure for different type of k^{th} Mul-HOI. These HOTrMs can achieve research Aims 3 and 4.

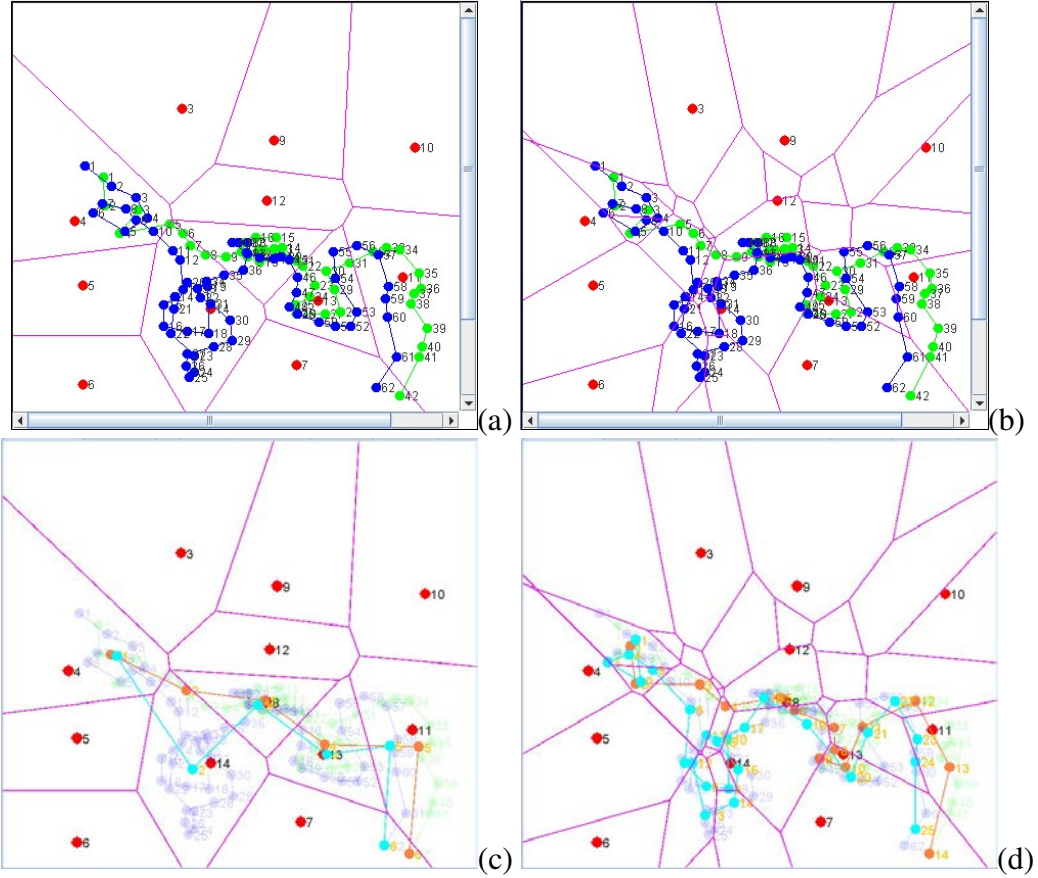


Fig. 5.12 OkVDs: (a) O1VD; (b) O5VD; topologically simplified: (c) O1VD; (d) O5VD with trajectory datasets $T_{21} = \{t_{21.1}, \dots, t_{21.42}\}$ and $T_{22} = \{t_{22.1}, \dots, t_{22.62}\}$ and generators $P_S = \{p_{S.1}, \dots, p_{S.12}\}$ (labelled from 3 to 14).

Figure 5.12(a) and 5.12(b) illustrate trajectory datasets T_{21} and T_{22} and a set P_S of generators. T_{21} is represented in green and T_{22} is represented in blue. Topologically simplified trajectory dataset T_{21} and T_{22} are shown in Figure 5.12(c) and 5.12(d). T_{21} is represented in orange and T_{22} is represented in light-blue. In this section, we present various treemap layouts for visualising topologically simplified trajectories. For HOI treemap calculation, the area of HOVD region for each trajectory node in a given trajectory is used to compute its corresponding sub-rectangle size in the treemap, which is similar to the original treemap calculation. In this study, a VA tool is designed to visualise two treemaps for two trajectories

T_{21} and T_{22} : boundaries of rectangles for T_{21} in green on left and boundaries of rectangles for T_{22} in blue on right. Areas of rectangles are coloured automatically from our treemap package. T_{21} and T_{22} shown in Figure 5.12 will be used for the following treemap layouts.

5.4.1 Binary Treemaps for HOI

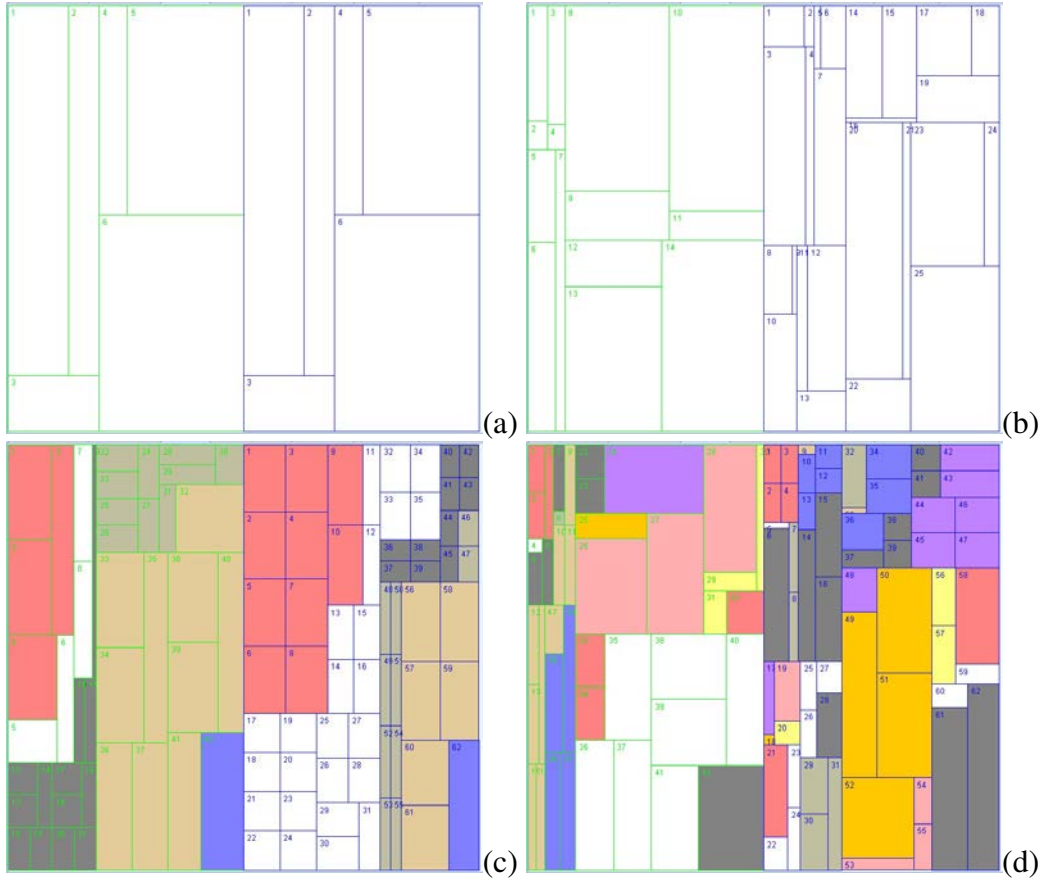


Fig. 5.13 Binary treemaps for HOI: (a) O1VD with topologically simplified T_{21} and T_{22} ; (b) O5VD with topologically simplified with T_{21} and T_{22} ; (c) O1VD with T_{21} and T_{22} ; (d) O5VD with T_{21} and T_{22} .

The binary tree tiling should be considered when the layout stability is extremely important. While not having shapes with very good aspect ratios, it is very stable when changes occur in the data.

Figure 5.13(a)-5.13(d) show the simplest and straightforward mapping of binary tree to a treemap with original trajectory HOI and topologically simplified HOI for trajectory datasets

T_{21} and T_{22} . Note that from Figure 5.13(a), we can see that treemaps of the topologically simplified HOI of O1VD for both T_{21} and T_{22} are exactly the same as those shown in Figure 5.13(a). However, the higher-level HOI reveals differences as shown in Figure 5.13(b). Similarly, Figure 5.13(c)-5.13(d) combine both the geometrical (node) and topological (colour) HOI in the same figure. Geometrical HOI is used to obtain sub-rectangle sizes of the treemaps and data points with the same topological HOI coded in the same colour. Figure 5.13(a) shows that two trajectories have the same topologically simplified HOI of O1VD. However, Figure 5.13(c) has different binary treemap for original trajectory HOI. Although those two trajectories display different binary treemaps for HOI, area colour shows that those areas have the same topologically simplified relations.

5.4.2 Squarified Treemaps for HOIs

A squarified tree is probably the best general purpose layout algorithm as it seeks a tiling where all rectangles have an aspect ratio as close to one as possible (i.e., a square). [113] addresses the problem of elongated rectangles resulting from the simple subdivision technique of binary treemaps. It is very difficult to identify trajectory points (nodes) with small values in standard binary treemaps. Squarified treemaps are more space-efficient as the number of pixels to be used for the border is proportional to the circumference, and this proportion is minimal for a square. Also, squares are easier for users to compare elongated rectangles.

Figure 5.14 presents squarified treemaps for HOI using the algorithm described in [113]. Figure 5.14(a) shows two trajectories T_{21} and T_{22} have the same topologically simplified HOI, when $k = 5$, it displays difference between two trajectories. However, when nodes have been changed, the squarified treemaps can have a completely different treemaps structure. It is displayed in more detail in Figure 5.14(c)-5.14(d).

5.4.3 Strip Treemaps for HOIs

This is a variation of the squarified tree layout algorithm, and it differs because the layout direction is fixed from the start. One of its key strengths, other than its natural tiling that

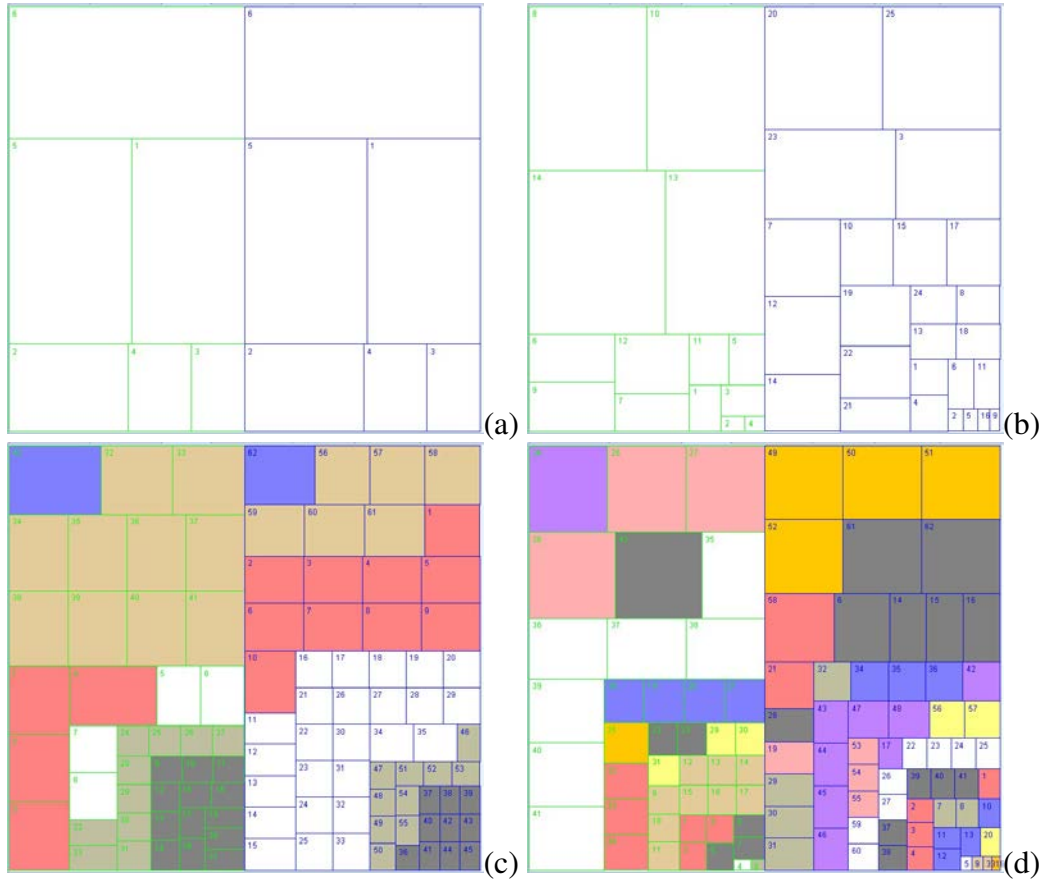


Fig. 5.14 Squarified treemaps of for HOI: (a) O1VD with topologically simplified T_{21} and T_{22} ; (b) O5VD with topologically simplified with T_{21} and T_{22} ; (c) O1VD with T_{21} and T_{22} ; (d) O5VD with T_{21} and T_{22} .

produces excellent readability, is that it is ordered and can therefore have its content organised in a predictable way. It processes input data areas in order and displays the sub-rectangles in horizontal or vertical strips of varying thickness. It is more computationally efficient than squarified treemaps because it only looks at areas within one strip, which is currently being processed.

Figure 5.15 presents strip treemaps for HOI. Figure 5.15(a)-5.15(b) show topologically simplified HOI when $k = (1, 5)$. It has similar but different treemaps information compared with squarified treemaps in Figure 5.14. Figure 5.15(c)-5.15(d) display T_{21} and T_{22} original trajectories HOI with topologically simplified HOI (colour).

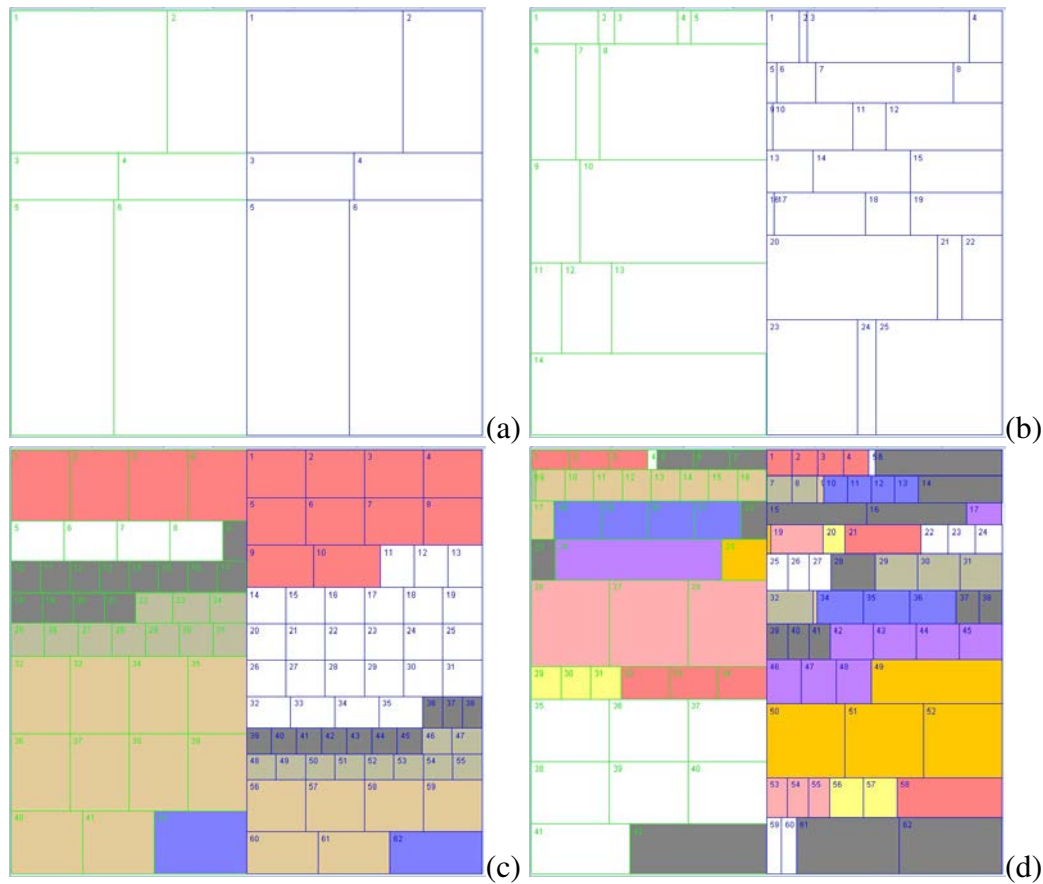


Fig. 5.15 Strip treemaps for HOI: (a) O1VD with topologically simplified T_{21} and T_{22} ; (b) O5VD with topologically simplified with T_{21} and T_{22} ; (c) O1VD with T_{21} and T_{22} ; (d) O5VD with T_{21} and T_{22} .

5.4.4 Slice Treemaps for HOIs

Invented in 1991, this was the first treemap layout algorithm. Its biggest problem is poor aspect ratios; the tiling becomes less useful for large datasets or screen real estate. For each level, the layout direction is changed between horizontal and vertical, and the display is proportionally divided among the children. Orders of data points are retained in the visualisation.

Figure 5.16(a)-5.16(b) show slice treemaps of topologically simplified HOI, and Figure 5.16(c)-5.16(d) display geometrical HOI of T_{21} and T_{22} . Interestingly, these parallel thin rectangles are quite visually distinctive for identifying clusters, such as data points with the same topological HOI.

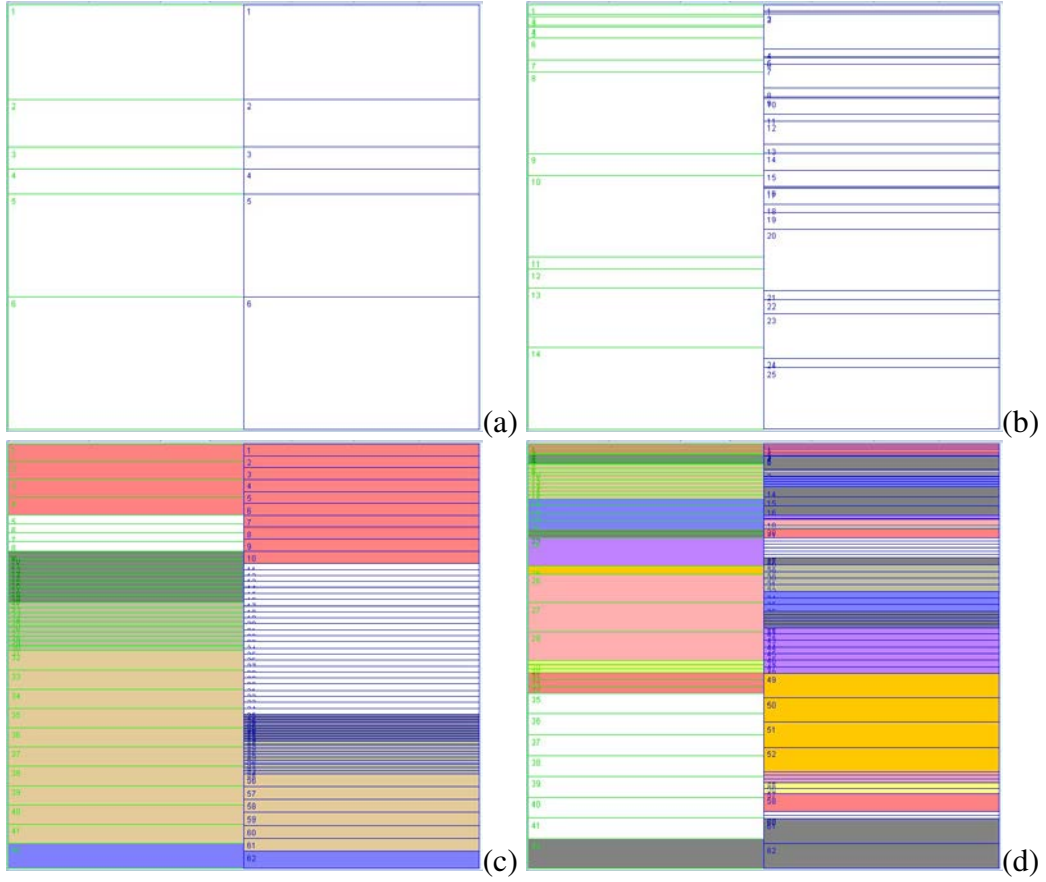


Fig. 5.16 Slice treemaps for HOI: (a) O1VD with topologically simplified T_{21} and T_{22} ; (b) O5VD with topologically simplified with T_{21} and T_{22} ; (c) O1VD with T_{21} and T_{22} ; (d) O5VD with T_{21} and T_{22} .

5.4.5 Ordered Treemaps for HOIs

An ordered treemap creates a treemap layout of data points that are next to each other in a given order and are put next to each other in the treemap. Thus, ordered treemap is a good choice in situations where legibility, usability, and smooth updating all are important concerns.

Figure 5.17 presents corresponding ordered treemaps. Since trajectory datasets contain the temporal dimension, the ordered treemaps can be used to display the dynamics of moving objects. For example, from Figure 5.17(a) and 5.17(b), we can see the changes of topological HOI along the sequence of the data points in T_{21} and T_{22} .

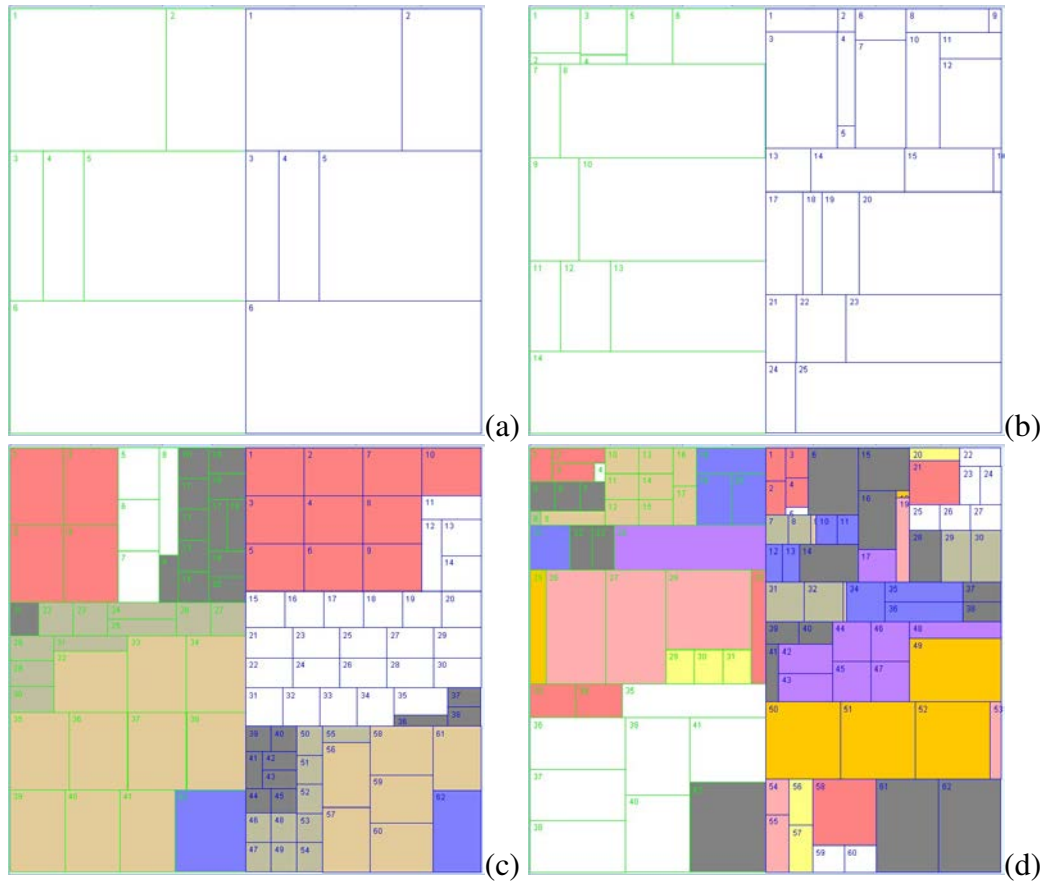


Fig. 5.17 Ordered treemaps for HOI: (a) O1VD with topologically simplified T_{21} and T_{22} ; (b) O5VD with topologically simplified with T_{21} and T_{22} ; (c) O1VD with T_{21} and T_{22} ; (d) O5VD with T_{21} and T_{22} .

5.4.6 Pivot by Size Treemaps for HOIs

Pivot by size is a variation of ordered treemaps [114]. A pivot, that is the data point with the largest area in the list, is chosen and will be put first in the visualisation because it is the most difficult to arrange in the overall visualisation and also has the most impact to the overall layout.

Figure 5.18(a) and 5.18(b) display pivot by size treemaps of topologically simplified trajectory data T_{21} and T_{22} and Figure 5.18(c) and 5.18(d) show original trajectory data of geometrical HOI.

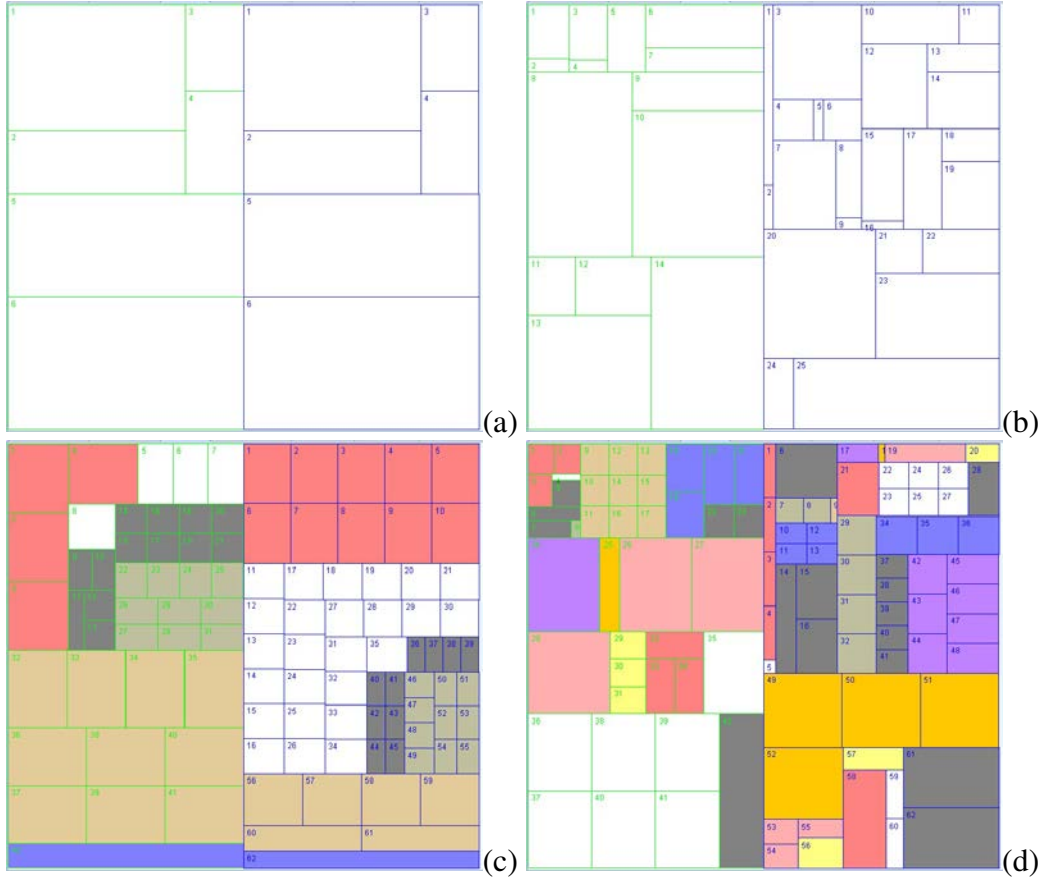


Fig. 5.18 Pivot by size treemaps for HOI: (a) O1VD with topologically simplified T_{21} and T_{22} ; (b) O5VD with topologically simplified with T_{21} and T_{22} ; (c) O1VD with T_{21} and T_{22} ; (d) O5VD with T_{21} and T_{22} .

5.4.7 Pivot by Split Size Treemaps for HOIs

Pivot by split size treemaps use the data point which will split the remaining data points into approximately equal total areas as the pivot. A more balanced layout can be produced by this approach. Being a partially ordered algorithm, the readability is better than the squarified treemap although not as good as the strip layout, which follows a simple linear pattern.

Figure 5.19 presents corresponding pivot by split size treemaps. Figure 5.19(a)-(b) show topologically simplified HOI when $k = (1, 5)$. Figure 5.19(c)-(d) display original geometrical HOI with O1VD and O5VD.

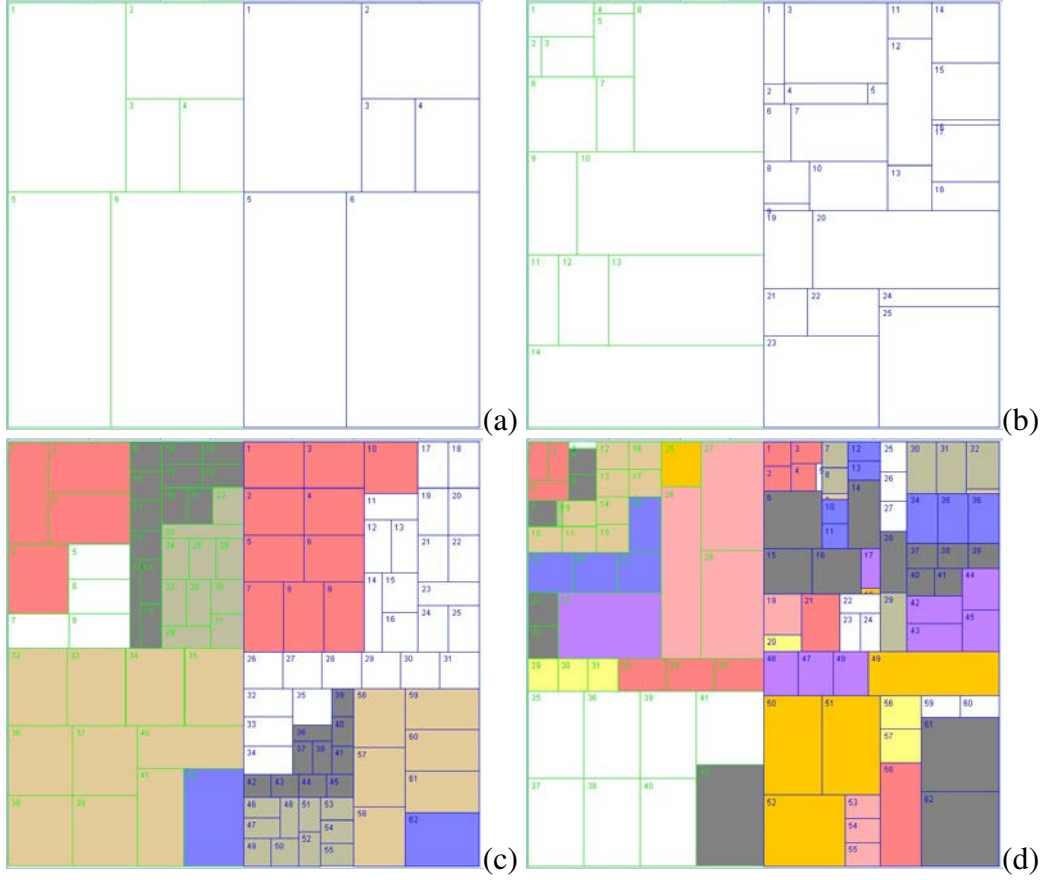


Fig. 5.19 Pivot by split size treemap for HOI: (a) O1VD with topologically simplified T_{21} and T_{22} ; (b) O5VD with topologically simplified with T_{21} and T_{22} ; (c) O1VD with T_{21} and T_{22} ; (d) O5VD with T_{21} and T_{22} .

5.4.8 Treemap Comparison

In order to compare different treemap layout algorithms, we use the following two metric measures: 1) the average aspect ratio of treemaps layout; 2) a layout distance change function which quantifies how much rectangles move as data are updated and established in [114]. A better design should have a lower average aspect ratio and a low distance change when data is updated. We adopt the unweighted arithmetic average of the aspect ratios of all sub-rectangles for each treemap for each layout and average overall of all treemaps of all level of HOI for each trajectory in each layout. In Table 5.4, the left part shows the aspect ratio of various treemaps layouts. For the layout distance change function, we measure how much the layouts are different from each level of HOI of the geometrical HOI. For

5.4 Higher-Order Treemaps (HOTrM)

Table 5.4 Comparison of treemaps in terms of aspect ratio and distance change.

Treemap	Aspect Ratio		Average	Changes	Average
Binary	Geo	T_{21}	8.8	T_{21}	111.35
		T_{22}	3.52		
	Top	T_{21}	19.19	T_{22}	92.15
		T_{22}	7.45		
Ordered	Geo	T_{21}	2.41	T_{21}	131.21
		T_{22}	2.11		
	Top	T_{21}	5.67	T_{22}	130.23
		T_{22}	3.24		
PBS	Geo	T_{21}	2.97	T_{21}	158.24
		T_{22}	2.52		
	Top	T_{21}	4.23	T_{22}	159.85
		T_{22}	2.98		
PBSS	Geo	T_{21}	2.04	T_{21}	150.58
		T_{22}	1.9		
	Top	T_{21}	3.37	T_{22}	155.64
		T_{22}	2.81		
Slice	Geo	T_{21}	83.52	T_{21}	57.19
		T_{22}	95.07		
	Top	T_{21}	30.61	T_{22}	52
		T_{22}	34.63		
Squarified	Geo	T_{21}	1.28	T_{21}	232.93
		T_{22}	1.24		
	Top	T_{21}	1.88	T_{22}	230.27
		T_{22}	1.52		
Strip	Geo	T_{21}	1.91	T_{21}	165.66
		T_{22}	1.83		
	Top	T_{21}	3.44	T_{22}	153.7
		T_{22}	2.53		

instance, measuring how the treemaps layout of O1VD is different from O2VD, then O2VD to O3VD, and up to O11VD. The distance change measure for topological HOI will be irrelevant because each level has a different number of sub-rectangles. The results are shown in the right part in Table 5.4. None of the layouts excelled in both aspect ratios and distance changes. A lower aspect ratio allows users to identify and to select the sub-rectangle in the treemap more easily and a low distance change will be required for users to keep track of the

data points when the data is updated. Hence, a treemap layout should be chosen according to the purpose of data analysis and applications. Also, these measures may be sensitive to the nature and value of the dataset used. More testing with different types of trajectory datasets will be required to understand the effectiveness and efficiency of each layout for VA of HOI trajectories. The proposed VA tool enables users to explore these different treemaps for context-dependent decision-making.

5.4.9 Summary and Discussion

In this section, we consider trajectory datasets where each trajectory is represented by a set of point sequences. Hence, HOI (k NN and k OR) is based on the Euclidean distance between points in trajectories. This section proposes a visual analytical approach to explore and compare geometrical and topologically simplified HOI of trajectory datasets. Previous research has revealed no study in this area about which treemap supports multiple moving objects of HOI. This approach enables users to explore insights of multiple trajectories with regard to POIs. The tool implemented for this approach uses the unified Delaunay data structure to support a treemap method and visualises HOI with 7 different types of treemap layouts. We show that this approach can help users explore similarities and differences of HOI for multiple trajectory datasets. Also, the tool supports the visualisation of both geometrical and topologically simplified trajectories HOI for the same location.

Pivot by split size treemap produces better results for all datasets but not the best one. The proposed tool enables users to interactively explore insights of a given dataset. The tool includes 7 types of treemap layouts. It allows users to steer the analysis process and choose the best layout for the data they are exploring.

The proposed approach is based on the unified Delaunay triangle data structure, thus it can effectively and efficiently support “what-if” analysis and interactive comparison of trajectory datasets. Future study includes inclusion of additional treemap layouts and related visualisation, such as tree-cone and 3-D treemaps to facilitate different VA and data mining.

Table 5.5 shows the contribution of methodology HOTrM to the research gap. HOTrM method fills the research gap of geometrical and topological HOI.

5.5 Higher-Order DNA (HODNA)

Table 5.5 Research contribution for HOTrM.

	Univariate/ Multivariate	Trajectory Support	Interactivity	Quantitative/ Qualitative	Higher-Order Information (HOI)			Higher-Order Dissimilarity Measurement			Higher-Order Data Mining
					Geo	Top	Dir	Geo	Top	Dir	
HOTrM	Uni/Mul	Yes	Yes	Quan/Qual	Yes	Yes	No	No	No	No	No

5.5 Higher-Order DNA (HODNA)

Original DNA Chart: DNA or Deoxyribonucleic acid is a type of nucleic acid. It is present in all living cells of bacteria, trees, dogs, cats and humans. Some viruses also contain DNA. DNA [115] was discovered in 1868 by the German biochemist Friedrich Miescher who called it nuclein. Nucleic acids are made up of chains of many repeating units called nucleotides. The DNA molecule actually consists of two such chains that spiral around an imaginary axis to form a double helix (spiral). The original DNA chart is a graphic to display the structure of genetic instructions and identify the unique colour-code by element. There is no research presenting a unique colour-code for moving objects of HOI. There is a need to find the unique and suitable colour-code to identify the dataset and compare a moving object's impact factor for measuring relationships between datasets.

Proposed HODNA: The proposed DNA chart will encode each trajectory with an unique colour sequence. By examining the colour sequences of different trajectories, topological HOI of these trajectories can be compared for identifying the common area where the trajectories moved to at the same time. Moreover, a measurement metric is proposed to measure topologically the impact factor for trajectories of k NN information. The proposed HODNA can achieve research Aim 4 and measurement metric can achieve research Aim 6.

5.5.1 Higher-Order DNA Chart (HODNAC)

In this section, we present a visualisation approach for displaying topological HOI called Higher-Order DNA Chart (HODNAC). HODNAC visualises relationship of topological HOI in: (1) trajectory datasets of each k^{th} POIs; (2) k^{th} HOI relationship between each trajectory with yes/no.

To create a HODNAC, first of all, topological information for each trajectory node $t_k \in T$ (original trajectory T for topological HODNAC) of kNN is calculated for that ID of POIs. Once the ID of each trajectory node is calculated then it is compared with previous or next trajectory node as normalised into [yes/no] for relative visualisation. Then, topological information for each trajectory node $d_k \in DNA$ (original trajectory DNA for topological HODNAC) is determined for that topology from POIs.

Figure 5.20 shows Voronoi diagram and HODNAC of two different trajectories datasets T_{23} and T_{24} and generators P_T . Figure 5.20(a) presents trajectory T_{23} in green with generators P_T in red, Figure 5.20(b) depicts a HODNAC for trajectory dataset and generator from Figure 5.20(a). Figure 5.20(c) illustrates trajectory T_{24} in green with generators P_T in red, Figure 5.20(d) presents a HODNAC for trajectory dataset and generator from Figure 5.20(c). The DNA chart consists of each order k of different layers showing trajectory nodes with different kNN . The longitude shows node ID of trajectory dataset movement over time from left to right. The latitude shows each k of HOI based on trajectory datasets. The diagram not only shows each trajectory node with different k^{th} POIs ID, but also displays each k^{th} HOI of trajectory datasets. The diagram also considers the topological information when each trajectory node has the same topological information (POIs' ID) with the previous or next trajectory node, it draws a relationship link line within red colour which is "YES" topological information. When a trajectory node has different topological information from the previous or next trajectory node, it draws a relationship link line in blue which is "NO" topological information.

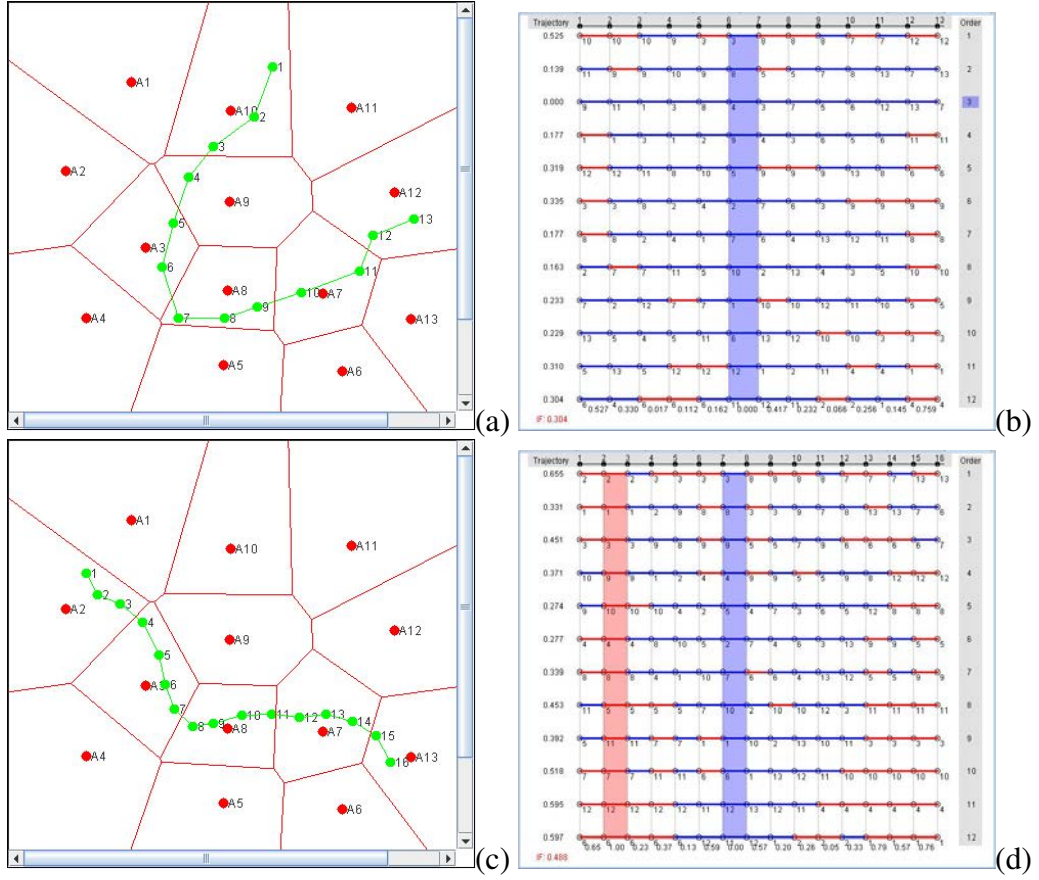


Fig. 5.20 OkVDs: (a) O1VD and (b) HODNAC for trajectory $T_{23} = \{t_{23.1}, \dots, t_{23.13}\}$ based on $P_T = \{p_{T.1}, \dots, p_{T.13}\}$ (labelled from A1 to A13); (c) O1VD and (d) HODNAC for trajectory $T_{24} = \{t_{24.1}, \dots, t_{24.16}\}$ based on $P_T = \{p_{T.1}, \dots, p_{T.13}\}$ (labelled from A1 to A13).

5.5.2 Higher-Order Trajectory Area (HOTA)

In this section, we introduce a visualisation approach for presenting HODNAC called Higher-Order Trajectory Area (HOTA). HOTA visualises geometrical trajectory area with: (1) the area shape that has the same topological (Order-1 to Order- k) HOI of each trajectory dataset; (2) area's relationship between trajectories with common area sharing; (3) geometrical information of trajectory area visualisation based on topological HOI of HODNAC.

To build a HOTA, first of all, topological information for each trajectory node $t_k \in T$ (original trajectory T for geometrical HOTA) of k NN is calculated for that area of POIs. Once the area of each trajectory node is calculated, then combine them all together as normalised into a common area for relative visualisation. Then, the area (geometrical information)

for each trajectory node $ta_k \in HOTA$ (original trajectory area for geometrical HOTA) is determined for that geometry between POIs and trajectory.

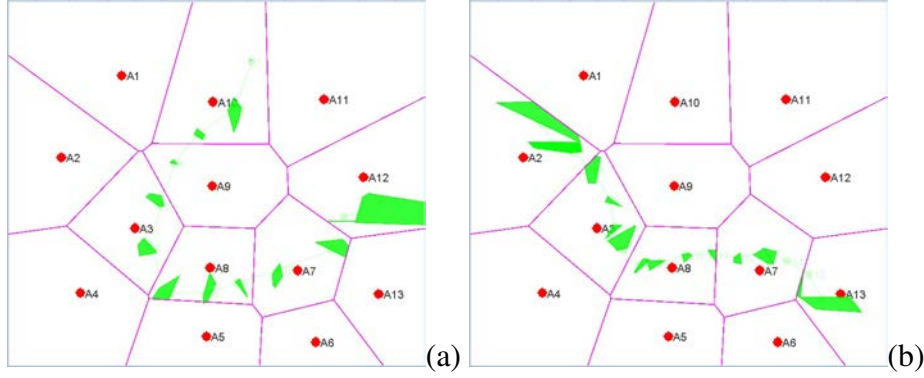


Fig. 5.21 HOTA: (a) HOTA for trajectory T_{23} based on P_T . (b) HOTA for trajectory T_{24} based on P_T .

Figure 5.21 shows HOTA of two different trajectories datasets and POIs shown in Figure 5.20. The HOTA consists of each area based on trajectory nodes of kNN information. It draws the geometrical area information within the same topological HOI of trajectory from HODNAC.

5.5.3 Higher-Order DNA Impact Factor (HODNAIF)

In order to justify top- k HOI for higher-order trajectory and relationship between trajectory datasets and POIs, we introduce a new measurement approach called Higher-Order DNA Impact Factor (HODNAIF). HODNAIF is a mean approach to calculate an average factor between the trajectory nodes' impact factor and up to higher-order k^{th} impact factor.

“Mean” method is used synonymously to refer to a central value of a discrete set of numbers, specifically, the sum of the values divided by the number of values. For trajectory impact factor, we use “mean” method to calculate topological HOI. First, each trajectory node finds k^{th} order POIs; then, Trajectory Single Factor (TSF) calculates all k^{th} order POIs “Mean” with previous or next trajectory node to find out the mean of two trajectory nodes HOI.

The Trajectory Factor (TF) is defined as below:

$$TSF_i = \sum_{a=1}^i \sum_{b=1}^k (1 - \frac{d_a}{d_b}). \quad (5.11)$$

$$TF_i = \sum_{c=1}^k (\frac{TSF_i}{TSF_c}). \quad (5.12)$$

For example, given a trajectory $T = \{n_1, \dots, n_t\}$ of t nodes and a set $P = \{p_1, \dots, p_m\}$ of m generators, let d_i return a subset $P_m \subset P$ of k^{th} order nearest neighbours, and find the “distance” of d_i and sum of all nearest neighbours $\sum_{j=1}^k d_j$, and find the new TF_i for k^{th} factor. In the end, the Root Mean Square (RMS) for all trajectory nodes of order k is computed when they have the same topological information TOP_{TF_i} . The Trajectory Impact Factor (*TIF*) is defined as the square root of mean square distance between current trajectory node with previous trajectory node. Mathematically it is defined as:

$$\begin{cases} TIF_{[t,t-1]} = \sum_{i=2}^k \sqrt{\frac{1}{2}(TF_i^2 + TF_{i-1}^2)} & \text{if } TOP_{TF_i} == TOP_{TF_{i-1}}. \end{cases} \quad (5.13)$$

When *TIF* is 1, the area between trajectory nodes will be filled in red colour, which means these two trajectory nodes have the same topological information. On the other hand, when *TIF* is 0, the area between trajectory nodes will be filled in blue colour, which means these two trajectory nodes have completely different topological information. It is shown in Figure 5.20 (b) and (d). For example, in Figure 5.20 (b), the *TIF* for node $T_{23} - 2$ (T_{23} with order-2) and $T_{23} - 3$ (T_{23} with order-3) is 0.330. In Figure 5.20 (d), the trajectories $T_{23} - 2$ and $T_{23} - 3$ have exactly the same topological information as filled in red colour. The trajectory $T_{23} - 7$ and $T_{23} - 8$ has totally different topological information which has been filled in blue colour. In some case, *TIF* can include 1 and 0 at the same time, which means in trajectory dataset, some nodes will have the same topological information and some nodes will have totally different topological information.

For the higher-order k^{th} impact factor, first, each k^{th} HOI finds trajectory node POIs; then, it calculates all POIs “distance” $\sum_{m=1}^t d_m$ through trajectory node changes over time to find out the distance of k^{th} order trajectory nodes HOI. The Order Factor (*OF*) is defined as below:

$$OF_i = \sum_{m=1}^t \left(\frac{d_i}{d_m} \right). \quad (5.14)$$

For example, let d_i return a subset $P_m \subset P$ of k^{th} order nearest neighbours, and find the “distance” of d_i and sum of all trajectory nodes $\sum_{m=1}^t d_m$, and calculate the new OF_i for trajectory node factor. In the end, RMS for all trajectory nodes of order k^{th} is computed when they have the same topological information TOP_{OF_i} . The Order Impact Factor (OIF) is defined as the square root of mean square distance for all trajectory nodes k^{th} generators. Mathematically it is defined as:

$$\left\{ OIF_k = \sum_{t=2}^T \sqrt{\frac{1}{T-1} (OF_t^2 + OF_{t-1}^2)} \quad \text{if } TOP_{OF_t} == TOP_{OF_{t-1}}. \right. \quad (5.15)$$

When OIF is 1, the order k will highlight in red rectangle colour in the “Order” column, which means in this order k^{th} the trajectory dataset has the same topological information. On the other hand, when OIF is 0, which means the trajectory dataset has totally different topological information in order k^{th} , it will highlight in blue rectangle in the “Order” column. It is illustrated in Figure 5.20 (b) and (d). For example, in Figure 5.20 (b), the OIF for order 2^{nd} is 0.139. When order k is 3, it has totally different topological information when trajectory dataset changes over time. In some case, OIF can include 1 and 0 at the same time, which means in different order k , some order has the same topological information for trajectory dataset and some order has completely different topological information.

When a trajectory dataset is large, this method will calculate it by rounding up and down to the size that can fit into HODNAC. The relationship between TIF and OIF is defined as below:

$$\left\{ \begin{array}{ll} 0 < OIF \leq 1 & \text{if } TIF = 1, \\ 0 \leq OIF < 1 & \text{if } TIF = 0, \\ 0 < TIF \leq 1 & \text{if } OIF = 1, \\ 0 \leq TIF < 1 & \text{if } OIF = 0. \end{array} \right. \quad (5.16)$$

For HODNAIF, it calculates and justifies the general impact factor between trajectory datasets (TIF) and POIs (OIF). First, HODNAIF calculates all k^{th} order OIF and all trajectory nodes TIF by RMS to find out the mean of trajectory datasets and POIs. It displays at the bottom left of diagram in red colour. The HODNAIF is defined as below:

$$HODNAIF = \sqrt{\frac{1}{2} \left(\left(\frac{\sum_{i=1}^k OIF_i}{k} \right)^2 + \left(\frac{\sum_{t=1}^T TIF_t}{T-1} \right)^2 \right)}. \quad (5.17)$$

It is shown in Figure 5.20 (b) and (d). For example, in Figure 5.20 (b), the HODNAIF is 0.304.

5.5.4 Summary and Discussion

Nowadays, there are more and more different sources of data which contain valuable information that could be relevant for many different domains and researches. As the real world is turning into a big data environment, the big data explosion is now a reality and not just a buzz word, and it is hard to identify any area of research or practice which is not faced with the issues big data are associated with or the benefits that can be gained from it. By analysing this data, for example, we can manage and make our decisions smarter, predict demand or sentiment about our product or service with different locations, watch as trends in financial market set-up, or simply discover interesting patterns in our data we were not aware of.

HODNAC and HODNAIF are innovative methods which not only focus on highlighting topological relationships between trajectory dataset and POIs, but also allow decision-makers to compare different datasets for whether there are patterns worth analysing or big data worth researching. HODNAC can display with different k^{th} order information in the same chart. It filters geometrical and directional information from HOI and only displays topological information. For example, in emergency management, HODNAC can quickly help people find or point out changes of relationships with different plans (k^{th} order information), which do not need to consider geometrical or directional information at early stages of emergency management. It allows decision-makers to reach a decision quickly as timing is crucial in emergency management. HODNAIF is a new, innovative measurement

Visual Analytics Techniques

metric that can be calculated to compare the pattern differences between datasets. For example, in tourism research datasets, tourists' datasets can be collected in many different devices, like mobiles, computers, watches, cars, cameras, and so on. What kind of data is more useful for hospitals, hotels, and advertising companies; which datasets or dataset groups contain more patterns/relationships/informations (higher impact factor) than others from big data? HODNAIF provides a calculated result to compare and contrast different trajectory datasets and different POIs.

Table 5.6 Research contribution for HODNA.

	Univariate/ Multivariate	Trajectory Support	Interactivity	Quantitative/ Qualitative	Higher-Order Information (HOI)			Higher-Order Dissimilarity Measurement			Higher-Order Data Mining
					Geo	Top	Dir	Geo	Top	Dir	
HODNA	Uni/Mul	Yes	Yes	Quan/Qual	No	Yes	No	No	Yes	No	No

Table 5.6 shows the contribution of methodology HODNA that can be able to answer the research gap. HODNA method is to be able to fill the research gap of topological HOI and innovate new topological dissimilarity measurement.

5.6 Conclusion

Table 5.7 Summary of proposed VA tools.

	Univariate/ Multivariate	Trajectory Support	Interactivity	Quantitative/ Qualitative	Higher-Order Information (HOI)			Higher-Order Dissimilarity Measurement			Higher-Order Data Mining
					Geo	Top	Dir	Geo	Top	Dir	
Gap	?	?	?	?	?	?	?	?	?	?	?
HOPC	Uni/Mul	Yes	Yes	Quan/Qual	Yes	Yes	Yes	Yes	No	No	No
HOS	Uni/Mul	Yes	Yes	Qual	No	No	Yes	No	No	Yes	No
HOR	Uni/Mul	Yes	Yes	Quan/Qual	Yes	Yes	Yes	No	No	No	Yes
HOTrM	Uni/Mul	Yes	Yes	Quan/Qual	Yes	Yes	No	No	No	No	No
HODNA	Uni/Mul	Yes	Yes	Quan/Qual	No	Yes	No	No	Yes	No	No

In conclusion of this chapter, we propose new visual analytical approaches to analyse trajectories with multivariate datasets. The 5 different visual analytics approaches demonstrate

that the contributions can fill the research gaps and address our research aims. Table 5.7 shows how the proposed VA tools can be utilised to achieve the aims in this research. They can support multivariate datasets and trajectory datasets interactively. They can provide not only quantitative information, but also qualitative information. HOPC provides a parallel coordinate based method for geometrical, topological and directional HOI. A comparison of trajectory geometrical HOI is introduced using a dissimilarity measurement. HOS is a spider chart based visual approach to focus on directional HOI and a directional dissimilarity is introduced to compare with trajectory datasets directional information. HOR displays a radar chart based approach for geometrical, topological and directional HOI. A higher-order data mining approach is innovated for grouping trajectories HOI. HOTrM provides a tree-maps based visual data structure for geometrical and topological HOI. HODNA creates DNA charts for topological HOI and a dissimilarity measurement is introduced for topological information of trajectory datasets.

Chapter 6

Evaluation

This chapter explains the evaluation of the efficiency of the tools proposed in this research and provides a comparison of proposed approaches.

6.1 Efficiency

This section analyses the efficiency and capabilities of our framework and compares it with other existing approaches. Figure 6.1 shows existing HOI approaches. Figure 6.1(a) and Figure 6.1(b) show a Web map segmentation approach that is based on the raster approach [61]. The approach is able to construct approximate HOVDs with three popular Minkowski metrics (Euclidean, Manhattan, and Maximum distance) and is able to visualise HOI information. However, this research suffers from several major weaknesses that cannot be used for “what-if” emergency management.

First, it is based on the raster approach and is not able to represent topological and directional HOI. Second, it needs to reconstruct HOVDs for every task which is not time-efficient for emergency management. Third, it does not compute exact HOVDs, but only compute approximate HOVDs using a distance transformation. Note that accurate information is crucial in emergency management. Fourth, it is for univariate dataset and not suitable for multivariate dataset. Fifth, it is a simple visualisation tool and not an interactive VA tool. Sixth, it does not support a moving trajectory for dynamic HOI analysis.

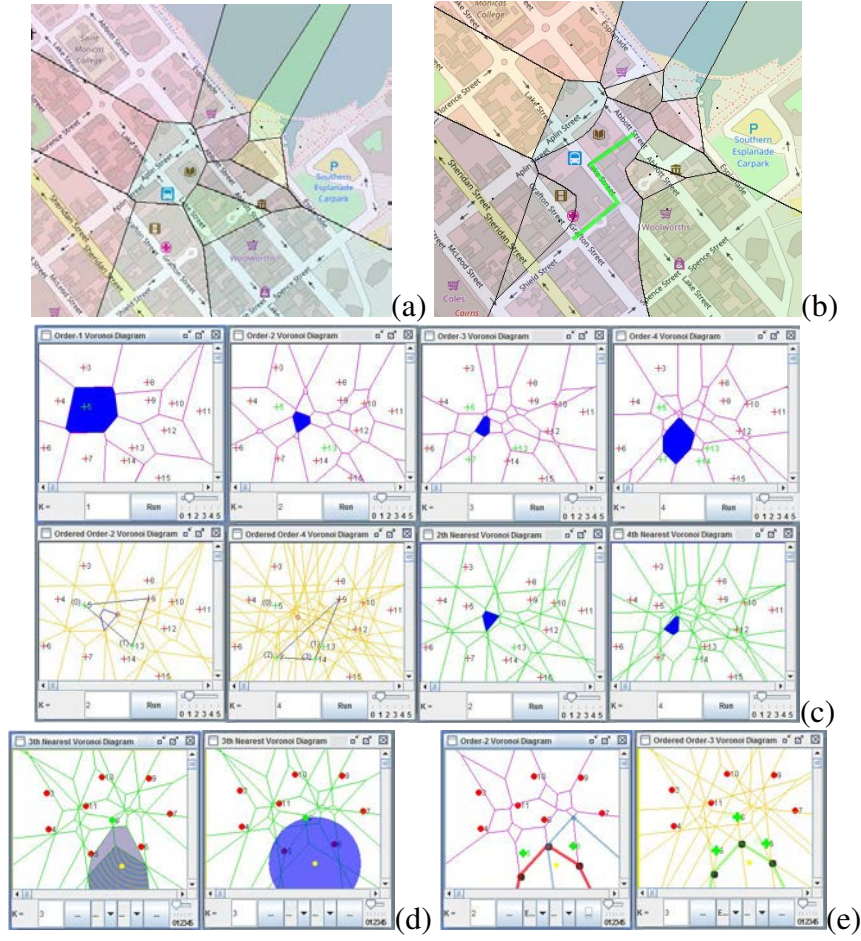


Fig. 6.1 Comparison of existing HOI approaches: (a) Raster-based for O1VD; (b) Raster-based for O1VD with a line; (c) HOVD families with kNN and kOR ; (d) Cushion and circle-region approach for Voronoi diagram; (e) Bevel approach for Voronoi diagram.

Figure 6.1(c) displays a complete set of HOVDs for $kNNQ$ and $kORQ$ [3]. It is based on the vector approach, and it supports a user to explore different k values for various HOVDs. It is able to support geometrical HOI and basic topological HOI, but unable to support directional HOI. As in [61], this approach only supports univariate datasets and fails to deal with a moving trajectory. In addition, this approach lacks high-level interactivity to enable user-oriented queries.

Figure 6.1(d) and Figure 6.1(e) illustrate various basic VA approaches for $kNNQ$ and $kORQ$ [4]. Voronoi areas (colours), edges (line thickness), and vertices (location) approaches enable a user to prune irrelevant, unnecessary information and to focus on interesting infor-

mation. The vector-based unified data structure not only provides a rich set of geometrical information but supports topological intra/inter/cross-relationships, thus they effectively and efficiently support “what-if” analysis and comparative analysis.

Table 6.1 Effectiveness and efficiency comparison with existing approaches (N : the number of grids; n : the number of POIs. Typically N is much larger than n).

	[61]	[3]	[4]	Proposed framework
Efficiency	$O(N^2)$	$O(n^3 \log n)$	$O(n^3 \log n)$	$O(n^3 \log n)$
Dataset	Univariate	Univariate	Univariate	Multivariate
Data structure	No	Yes	Yes	Yes
Interactivity	No	Low	Low	High
Trajectory Support	No	No	No	Yes
Geometrical HOI	Low	Low	Low	High
Topological HOI	No	Low	Low	High
Directional HOI	No	No	No	High
Dissimilarity Measure	No	No	No	High
HOI mining	No	No	No	High
HOVD reconstruction	Many	Once	Once	Once

Table 6.1 compares the effectiveness and efficiency of the proposed framework with existing studies [3, 61, 4]. The effectiveness and efficiency of the raster-based HOI approach [61] is totally dependent on the number of grid cells N and is unable to support many VA interactions. The vector-based HOI approach [3, 4] supports univariate datasets only, and does not deal with dynamic trajectories. In addition, it provides low functionality for various information analysis and very limited approaches are supported.

The raster-based approach HOI running time is $O(N^2)$ [61] that N is the number of grids. It considers single (univariate) dataset only without unified all order k data structure and this model cannot support trajectory dataset for k NN and k OR information, VA methods have non-interactivity with users, basic visualisation techniques have only focused on quantitative information, and information categories have mostly studied in geometrical information but cannot return topological and directional information. In addition, previous research does not provide any dissimilarity measure and HOI data mining techniques for each order k^{th} .

The vector-based approach HOI running time is $O(n^3 \log n)$ [3] where n is the number of POIs. It provides all order k unified data structure and supports univariate datasets only,

with very limited interactivity techniques and geometrical as well as topological information provided. There is no directional information support, and it does not deal with dynamic trajectories, which are two essential characteristics (multivariate and trajectory) in many applications such as in emergency management. Also, there is no unified data structure for dissimilarity measuring methods, and no HOI mining techniques that just need to be built once for HOVD reconstruction. In addition, this approach provides limited functionality and interactivity for various “what-if” emergency management processes. The basic VA HOI approach enables different ways for visualising k NN and k OR queries. Although it is similar to the vector-based HOI approach, it cannot support large sets of geometrical and topological information of intra/inter/cross relationships for k NN and k OR queries.

This proposed research framework overcomes all the drawbacks mentioned above, and proposes efficient VA tools for Mul-HOI. It provides not only a multivariate type of HOI (geometrical, topological and directional) at the same time, but also links together with dissimilarity measures and data mining methods to group trajectory datasets HOI comparisons.

6.2 Comparison and Contrast

This research proposes new visual analytical approaches to analyse trajectories with multivariate datasets for tourist management or emergency management. The unified data structure gives general suggestions or “what-if” information for managing tourism datasets or making plans in all stages of emergency management. The new multivariate data structure can support multivariate data analysis and can provide useful information for multivariate “what-if” analysis.

Table 6.2 shows the comparison of proposed VA tools. Our proposed VA tools provide users to interact with the program suite so that they can obtain three different types of HOI (geometrical, topological, directional) and get quantitative and qualitative HOI based on different POIs on the fly.

Table 6.2 relevant the contributions of different tools proposed in this thesis to the related research gaps. All 5 methodology research families answer research questions and research

Table 6.2 Comparison and contrast of proposed VA tools.

	Univariate/ Multivariate	Trajectory Support	Interactivity	Quantitative/ Qualitative	Higher-Order Information (HOI)			Higher-Order Dissimilarity Measurement			Higher-Order Data Mining
					Geo	Top	Dir	Geo	Top	Dir	
HOPC	Uni/Mul	Yes	Yes	Quan/Qual	Yes	Yes	Yes	Yes	No	No	No
HOS	Uni/Mul	Yes	Yes	Qual	No	No	Yes	No	No	Yes	No
HOR	Uni/Mul	Yes	Yes	Quan/Qual	Yes	Yes	Yes	No	No	No	Yes
HOTrM	Uni/Mul	Yes	Yes	Quan/Qual	Yes	Yes	No	No	No	No	No
HODNA	Uni/Mul	Yes	Yes	Quan/Qual	No	Yes	No	No	Yes	No	No

gaps mentioned in the previous section. They provide Mul-HOI, trajectory support and interactivity functionality. Note that, each individual approach fills certain gaps of this research as described as follows:

HOPC: quantitative and qualitative information support, providing parallel coordinate based geometrical types of HOI and introducing geometrical dissimilarity measurement methods for geometrical information comparisons between trajectory datasets.

HOS: qualitative information support, introducing new spider chart based visual methodology for directional HOI and directional dissimilarity measurement for trajectory datasets directional information.

HOR: qualitative and qualitative information support, providing radar chart based geometrical, topological and directional HOI, and introducing new HOI data mining methodology for grouping trajectory datasets HOI.

HOTrM: quantitative and qualitative information support, providing new tree-maps visual based geometrical and topological HOI for trajectory datasets.

HODNA: quantitative and qualitative information support, introducing new DNA chart based methodology for topological HOI and topological dissimilarity measurement for calculating the trajectories impact factor.

Chapter 7

Emergency Management Application

Emergency management is a discipline related to avoiding and dealing with both natural and man-made disasters. Decision-making can be a crucial difficulty throughout unexpected emergencies. Emergency management is composed of four basic phases: mitigation, preparedness, response, and recovery [116]. In this chapter, we demonstrate how the different VA tools can be used for real-world datasets. We explain the use of different methods and results with a study using geo-tagged photos from Flickr in Queensland, Australia, in the years 2010-2011. Flickr is a photo-sharing website. It has more than 6 billion images uploaded and shared by their users. For the Flickr dataset, there are 17,066 records preprocessed to produce the trajectory dataset used in this research. Figure 7.1 shows the distribution of the dataset using Google Earth 6.0 (earth.google.com). The aim of this chapter is to provide better insight into its usefulness in different scenarios.

Most of these geo-tagged records appear along the coastline of Queensland, where the major cities are located. This application will separately take a subset of the dataset with data points from Cairns, a regional city in north Queensland, Australia, to demonstrate the usefulness and applicability of the proposed visual analytical approaches.



Fig. 7.1 A study region and real world dataset: geo-tagged photos from Flickr.

7.1 HOPC for Emergency Management

This section demonstrates how the trajectory with Mul-HOI can be used for each stage of emergency management.



Fig. 7.2 A study region and real dataset: distribution of clusters in the Central Business District (CBD) of Cairns (P_U in red), and an information office in the CBD of Cairns (P_V in blue), and an emergency office in the CBD of Cairns (P_W in magenta).

Figure 7.2 shows a subset P_U of clusters (red) within the Cairns CBD. These locations can be set up for potential tourist centres related to emergency management. Blue datasets

P_V are current information offices within the Cairns CBD. Magenta datasets P_W are current emergency offices within the Cairns city.

7.1.1 Mitigation

Mitigation is a realisation that eliminates or decreases the causes of risk/disaster, and is a factor that provides prevention of (removing, eliminating, reducing) the effects of inescapable disaster.

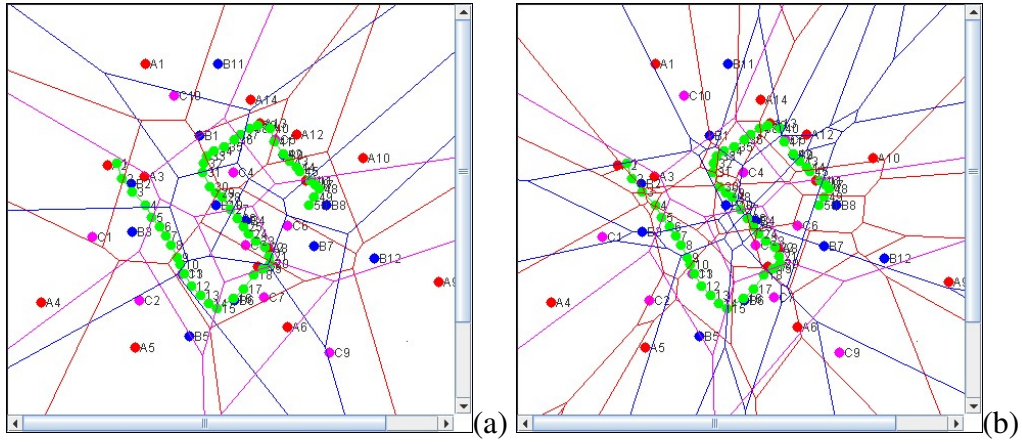


Fig. 7.3 (a) O1VD; (b) O5VD trajectory $T_{25} = \{t_{25.1}, \dots, t_{25.50}\}$ with Mul-HOI $P_U = \{p_{U.1}, \dots, p_{U.14}\}$ (labelled from A1 to A14), $P_V = \{p_{V.1}, \dots, p_{V.12}\}$ (labelled from B1 to B12), $P_W = \{p_{W.1}, \dots, p_{W.10}\}$ (labelled from C1 to C10).

Figure 7.3 shows different values of k with Order- k Voronoi diagrams. A trajectory dataset is collected every 30 seconds for 25 minutes from a tourist group which walked around the Cairns CBD. The data collected from tourist T_{25} is represented in green.

If we assume that P_U represents tourist attractions, P_V represents information centres, P_W represents emergency centres, Figure 7.3(a) shows O1VD, Figure 7.3(b) displays O5VD with only T_{25} data information. Considering the three POIs datasets are different, k is different for each. Based on Mul-HOI, it can provide useful information to indicate the nearest centres where people can seek help along the path of the trajectory, assuming that they can seek help from the emergency centres to come to the tourist attractions or the information centres.

7.1.2 Preparedness

In the preparedness phase, governments, organisations, and emergency centres make plans to save lives and minimise disaster damage such as having emergency personnel on standby or determining the nearest evacuation places with dynamic (moving/changing over time) data. In other words, preparedness also explores ways to improve damage-response operations.

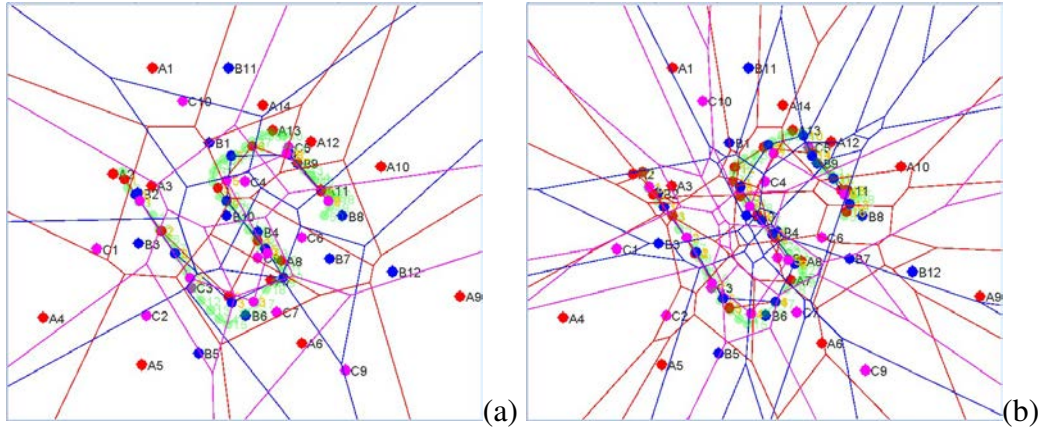


Fig. 7.4 OkVDs: (a) O1VD; (b) O5VD with 3 new topologically simplified trajectories dataset. T_U in red, T_V in blue, T_W in magenta.

Figure 7.4 shows different values of k with three new topologically simplified trajectory datasets. T_U is represented in red, T_V is represented in blue, and T_W is represented in magenta. The government can make decisions about the best location for setting up emergency tables or information tables for tourists before a disaster arrives. For the regions, the nearest neighbour emergency units give suggestions about areas which have more activities. These provide information for tourists to make back-up plans when the nearest information centre is not available or functional. With the additional information from the topologically simplified trajectory datasets, governments can make more informed decisions for planning emergency support before damage from the disaster occurs.

7.1.3 Response

Response consists of providing activities following the emergency situation to assist disaster victims. It also explores how to stabilise the situation and to prevent the probability of

secondary disaster and to give suggestions for recovery. Typically, the nearest neighbour emergency unit is required to respond to emergencies quickly. When emergencies occur, the nearest neighbour emergency units are required to cooperate and collaborate in order to decrease the damage from the disaster.

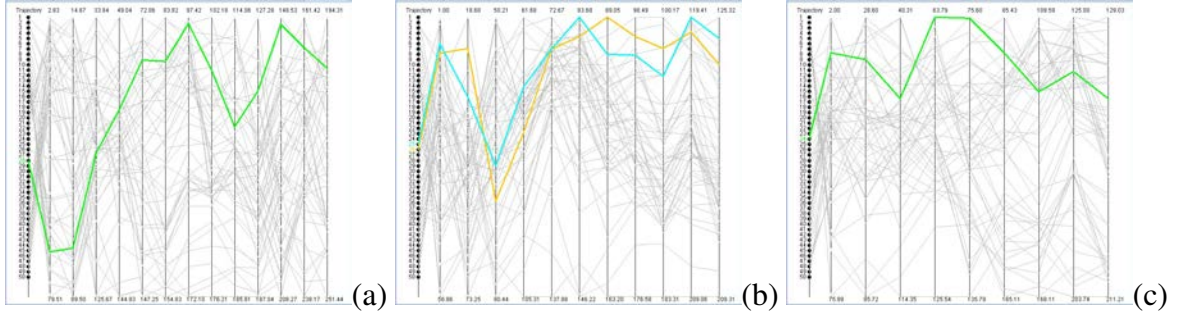


Fig. 7.5 PCBCs: (a) generator P_U 's PCBC; (b) generator P_V 's PCBC; (c) generator P_W 's PCBC.

Figure 7.5 shows different values of k with Order- k information. It displays information to determine the relationship of the trajectory based on generator P_U and P_V and P_W . It also highlights the quantitative PCMM and qualitative PCMM information between Mul-HOI trajectory T_{25} and generators P_U and P_V and P_W . The government focuses on these top-priority regions where the emergency centre and the information centre have the shortest distance to the trajectory, based on quantitative and qualitative PCMM. PCMM provides two ways to compare a set of POIs to trajectory Mul-HOI. It allows governments to make better decisions on where (location) and which (centre) begins to provide the emergency response.

7.1.4 Recovery

Recovery happens when the disaster is over. It is a set of activities that returns all systems back to normal and also prevents secondary damage. These include two phases, short term and long term. From Figure 7.6, we can see that although cluster numbers of topologically simplified trajectories are different, HOI is very similar. Figure 7.6(a) shows the common PCMM in red based on P_U . Figure 7.6(b) shows the common PCMM in blue based on P_V . Figure 7.6(a) shows the common PCMM in magenta based on P_W . In the short term of

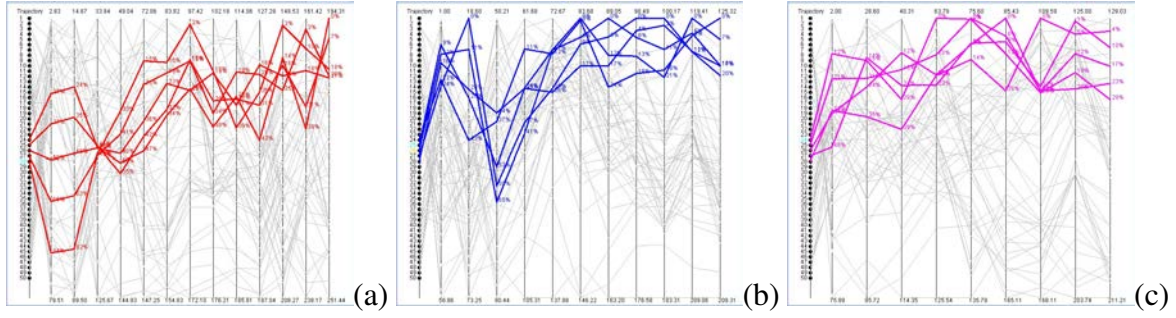


Fig. 7.6 Interaction PCBCs: (a) generator P_U 's PCBC; (b) generator P_V 's PCBC; (c) generator P_W 's PCBC.

emergency recovery, we can choose $T_{24} - T_{28}$ as recovery plans based on common quantitative and qualitative PCMM. For the long term, recovery plans can be introduced based on parallel coordinates pattern k of HOI. Interestingly, PCMM, based on those dataset are better PCMM after $k=5$, either trajectory with P_U or P_V or P_W . Therefore, for a long-term recovery plan, k after 5 of these datasets of HOI are the best references for decision-making.

7.1.5 Discussions

This study has proposed a new approach to analyse trajectory datasets for emergency management. The unified data structure gives general suggestions or information before making plans in mitigation procedures. Data structure can support with multivariate data information and prepare information for future analysis. Topologically simplified trajectory dataset can filter and well organise with different groups of people, to quickly make decisions in preparedness steps. People can logically obtain different information based on different POIs. In response steps, organisations or governments can support and control tourist attractions or information centres or emergency centres. Which, what, and how many items should be considered can be prepared for the next recovery step. An ability to make quick decisions for short-term recovery is important after disaster. This approach introduces short-term and long-term suggestions. Organisations can make decisions which cover multivariate centres. This is useful and will reduce disaster impacts.

7.2 HOS Application



Fig. 7.7 A study region and real dataset: distribution of clusters in CBD of Cairns P_U .

This application will take a subset of the dataset with data points from Cairns to demonstrate the usefulness and applicability of our proposed measures. Figure 7.7 shows a subset P_U of clusters within the Cairns CBD. These locations can be set up as information centres.

Figure 7.8 demonstrates HOVD of locations of 14 clusters. Figure 7.8 (a) shows O1VD with three photo-taker trajectories, while Figure 7.8 (b) displays O5VD. These clusters can be used to represent POI P_U . These three trajectory datasets in Figure 7.8 are collected every 10 seconds for half an hour from three different photo-taker groups walked to the Cairns lagoon (one of the most popular tourist destinations). The data collected from tourist T_{26} is represented in green, T_{27} is shown in blue, and T_{28} is depicted in light-pink in Figure 7.8. As P_U represents information centres, people moving around the study region will have different nearest information centres. That is, people passing through different paths will have a good understanding of which area has the same HOI when they move along.

Figure 7.8 shows a series of HOVD with original trajectories and also topologically simplified trajectories. T_{26} is represented in orange, T_{27} is in light-blue, and T_{28} is displayed in purple in Figure 7.8.

Figure 7.9(a) and Figure 7.9(b) show DSCs directional-geometrical HOI of the three trajectories. Figure 7.9(c) and Figure 7.9(d) display DSCs directional-topological HOI of the

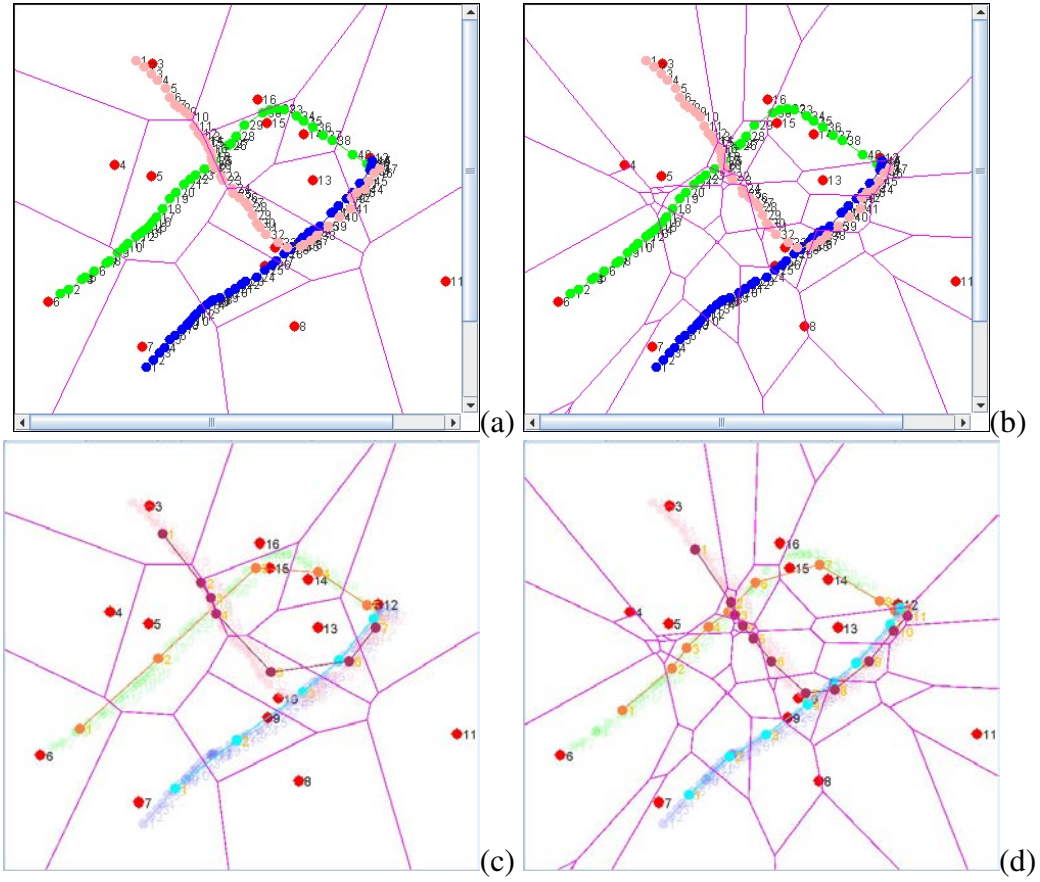


Fig. 7.8 $OkVD$ with trajectories: (a) O1VD; (b) O5VD with three trajectory datasets; (c) O1VD with topologically simplified trajectories; (d) O5VD with topologically simplified trajectories. Trajectory datasets T_{26} , T_{27} , and T_{28} with generators P_U .

three trajectories. T_{26} is represented in green, T_{27} is in blue, and T_{28} is shown in light-pink. Coloured areas in DSCs display relative relations of these three trajectories' directional HOI. In Figure 7.9(a), a DSC shows that the POIs are mostly located on the eastern side based on these three trajectories' O1VD information. However, when k increases to 5, Figure 7.9(b), directional information points out that most POIs stay on the western side. In Figure 7.9(c), a DSC shows the POIs are mostly located in eastern, northeastern, and southeastern sides, which is also evident in Figure 7.9(a). In Figure 7.9(d), a DSC depicts that *Southwest* is the most popular pointing direction, and also shows that *East* is another common direction.

Figure 7.10 shows DBRs of the three trajectory datasets. Interestingly, T_{26} and T_{28} exhibit a better directional balance than T_{27} for both directional-geometrical HOI and directional-topological HOI. Not surprisingly, directional-topological DBRs (Figure 7.10(b)) show simi-

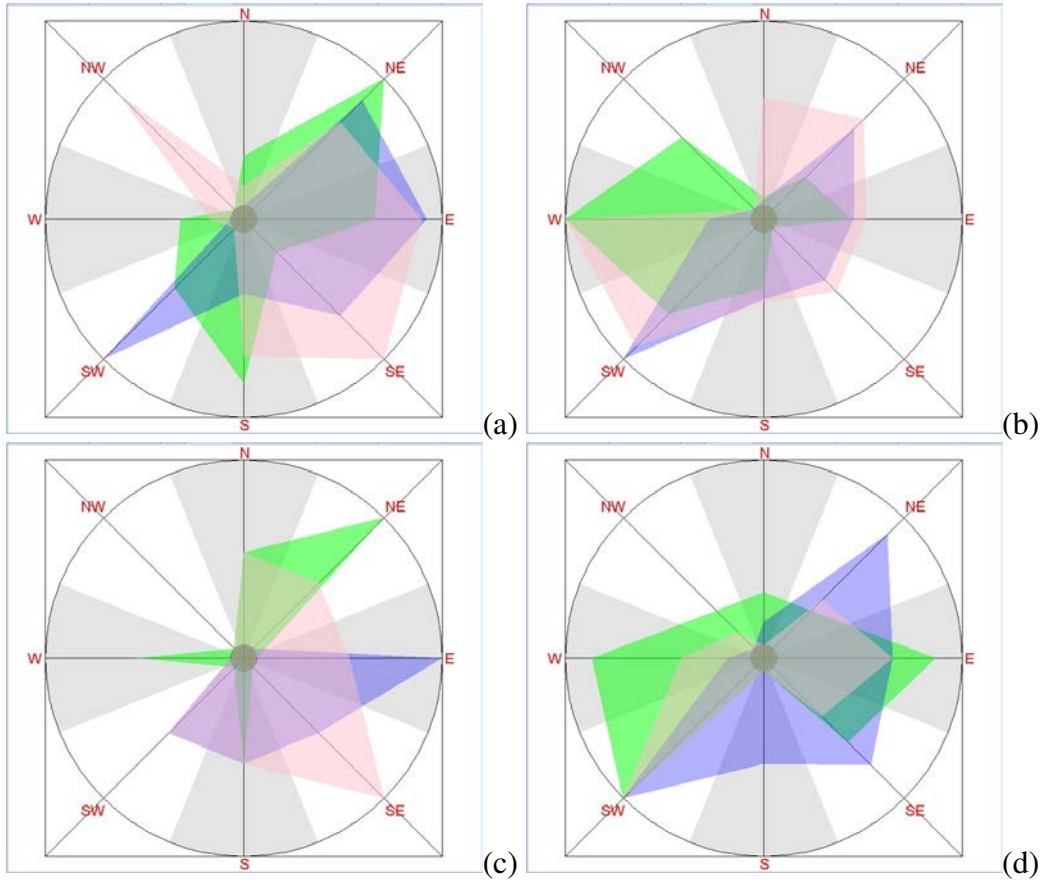


Fig. 7.9 DSCs: (a) O1VD directional-geometrical HOI; (b) O5VD Directional-geometrical HOI; (c) O1VD Directional-topological HOI; (d) O5VD Directional-topological HOI with datasets shown in Figure 7.8.

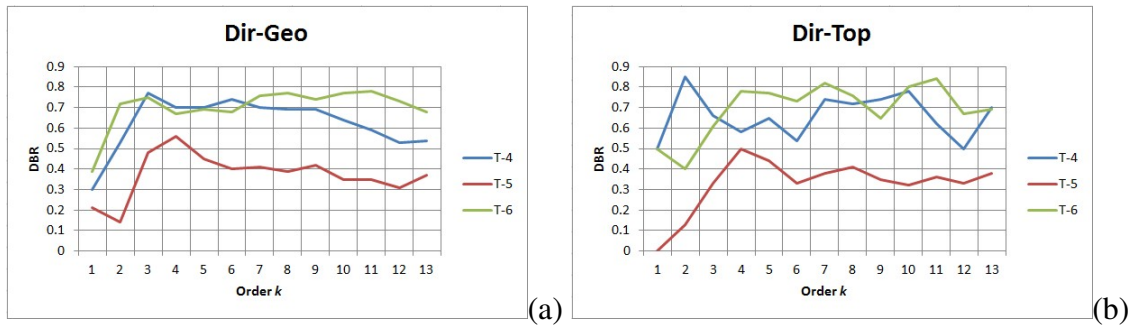


Fig. 7.10 Line graphs: (a) Directional-geometrical; (b) Directional-topological DBRs for datasets shown in Figure 7.8.

lar general patterns as directional-geometrical DBRs (Figure 7.10(a)). DBRs for directional-geometrical HOI for all three trajectories jump when k reaches to 3 and then form a plateau

for larger k values. This implies that for directional-topological HOI, these three datasets should have the 3^{rd} order HOI to have a more balanced directional distribution.

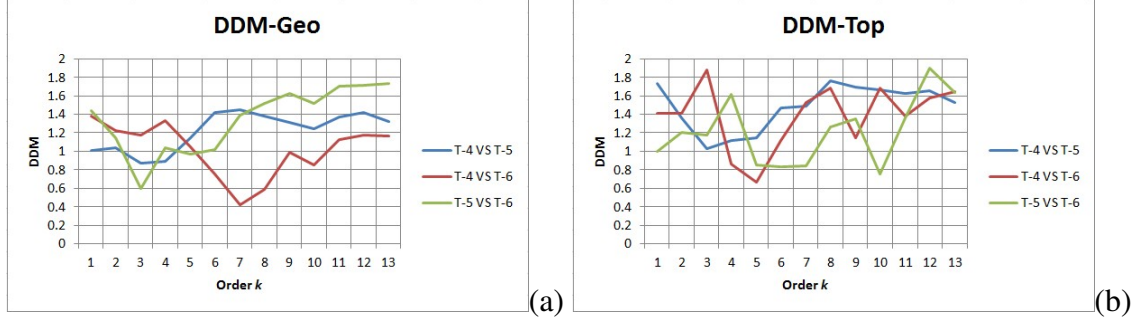


Fig. 7.11 Line graphs: (a) Directional-geometrical; (b) Directional-topological DDMs for datasets shown in Figure 7.8.

On one hand, Figure 7.11 shows that T_{26} and T_{28} have similar DDMs based on the original dataset (geometry). But in real dataset (Figure 7.8), these two datasets contain different geometrical information, including directions. In Figure 7.10, you can see those two trajectories have very high DBR, which means those two trajectories are quite well DBR. On the other hand, T_{26} and T_{28} ; T_{27} and T_{28} have very strong DDMs based on topologically simplified HOI (topology). In Figure 7.8, T_{26} and T_{27} have very similar geometry (movement), however, in Figure 7.10 DBR, directional information is totally different. In Figure 7.11, T_{26} and T_{27} have very high DDM values.

7.3 HOR Application

We illustrate how HORC and HO_k -Means can be used for emergency management using the dataset shown in Figure 7.7.

Figure 7.12(a) demonstrates HOVD of sites of 14 POIs P_U , and the O1VD with two photo-taker trajectories, while Figure 7.12(b) displays O5VD. P_U is a set of centroids determined for the Flickr dataset using k -means clustering of the photo records. The two trajectory datasets in Figure 7.12 are tourist movement datasets collected every 30 seconds for 15 minutes from two different tourists sightseeing in Cairns. The data collected from tourist T_{29} is represented in green and T_{30} is shown in blue in Figure 7.12. Let us assume P_U represents best locations

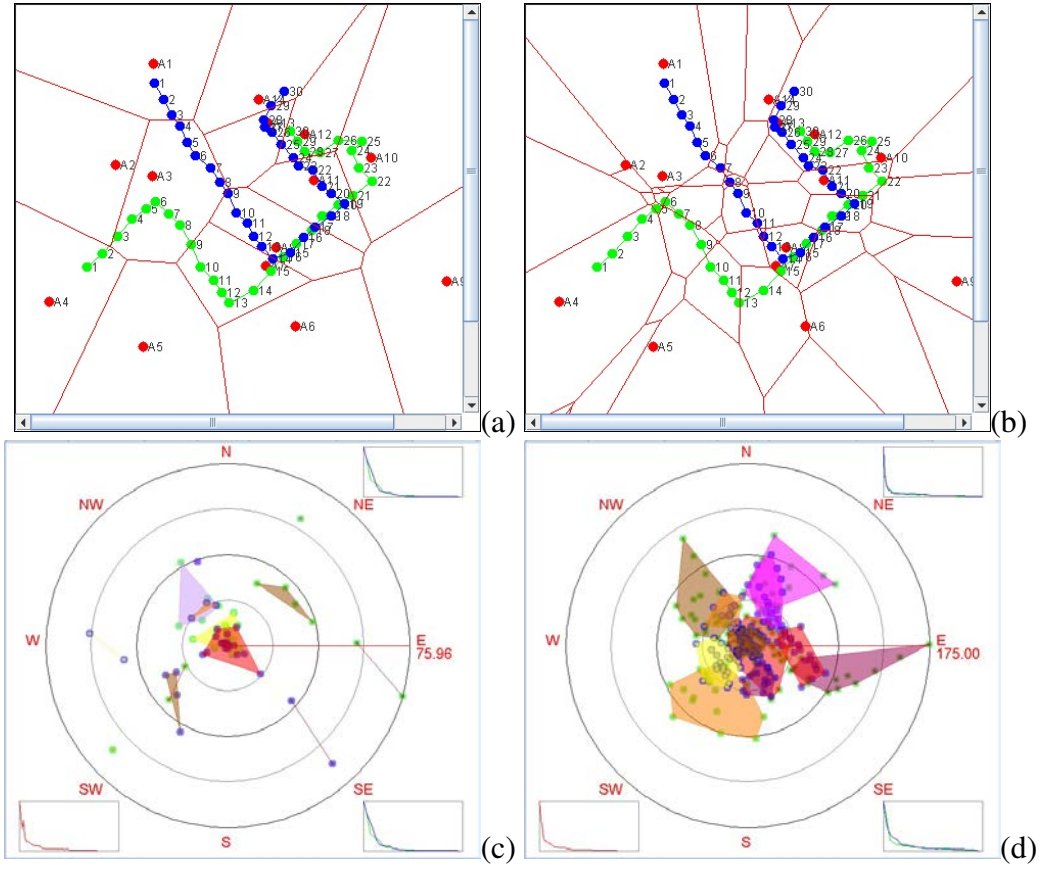


Fig. 7.12 (a) O1VD; (b) O5VD with two trajectory datasets; (c) HORIZ of O1VD (d) HORIZ of O5VD. Trajectory datasets T_{29} and T_{30} with generators P_U .

for setting up tourist information centres for tourists to seek help and advice when they are visiting the region.

Figure 7.12(c) and 7.12(d) show HORIZ and HOK-Means of the two trajectories. Figure 7.12(c) displays HORIZ in O1VD, T_{29} is presented in green, T_{30} is presented in blue. In Figure 7.12(c), a HORIZ shows that the POIs are mostly located on the west side and very close to a trajectory based on these two trajectories' O1VD information. However, when k increases to 5, as shown in Figure 7.12(d), T_{30} moves to the east and north. Obviously, most of the nearest POIs are located on the south-west side at around half of V_{Max} distance. T_{29} is located from west to east, and most of POIs are very close to the trajectory (less than 25% of V_{Max} distance).

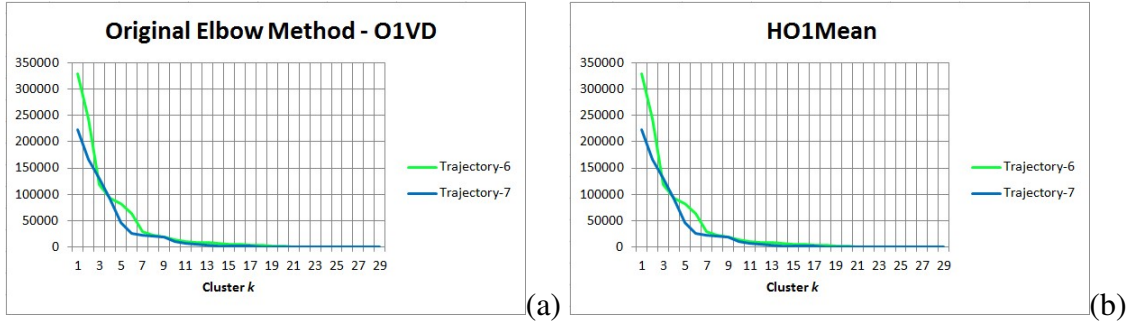


Fig. 7.13 (a) Original elbow method for O1VD; (b) HO1-Means of trajectory datasets T_{29} and T_{30} with generators P_U .

The clustering results for order-1 information are displayed in Figure 7.13(a) with an original elbow method, and in Figure 7.13(b) for the HO1-Means method. As we are only looking for the 1st nearest-neighbour HOI, the result will be the same because it calculates only one nearest-neighbour result. In Figure 7.13, we can see the elbow chart showing the SSE after running k -means clustering for k going from 1 to 29. We see a pretty clear elbow at $k=7$ for trajectory T_6 and $k=6$ for trajectory T_{30} , indicating that 6 is the best number of clusters for T_{29} and 7 is the best number of clusters for T_{30} .

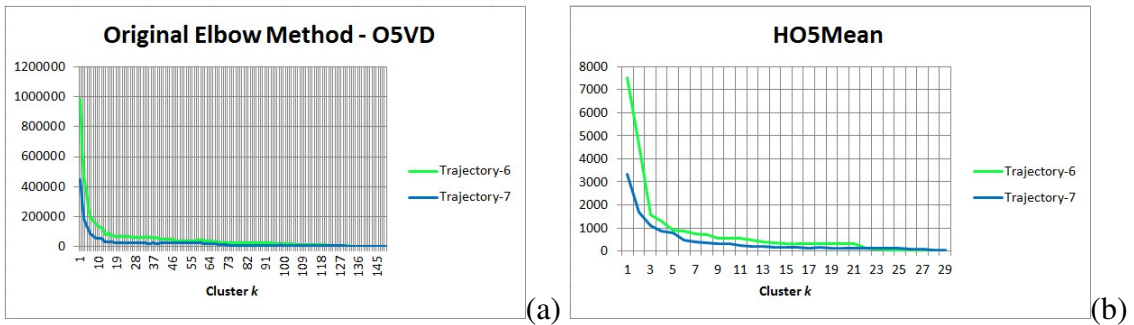


Fig. 7.14 (a) Original elbow method for O5VD; (b) HO5-Means of trajectory datasets T_{29} and T_{30} with generators P_U .

In another case, when we consider HOI for $k=5$, Figure 7.14(a) shows the original elbow method, and Figure 7.14(b) shows HO5-Means. Notice how the elbow chart for the original method does not have a clear elbow. It is a fairly smooth curve, and it is unclear what is the best value of cluster k . This result shows that the original method of k -means clustering does not work very well for trajectory HOI. On the other hand, Figure 7.14(b) presents the results by HO5-Means and suggests that the optimal number of clusters that the HO5-Means has

selected for T_{29} is $k=5$ and for T_{30} is $k=6$. Therefore, the proposed HO_k -Means method is superior to the original method for selecting the best k for clustering of trajectory HOI.

7.4 HOTrM Application

We illustrate how the treemaps of HOI can be used for emergency management using the dataset shown in Figure 7.7.

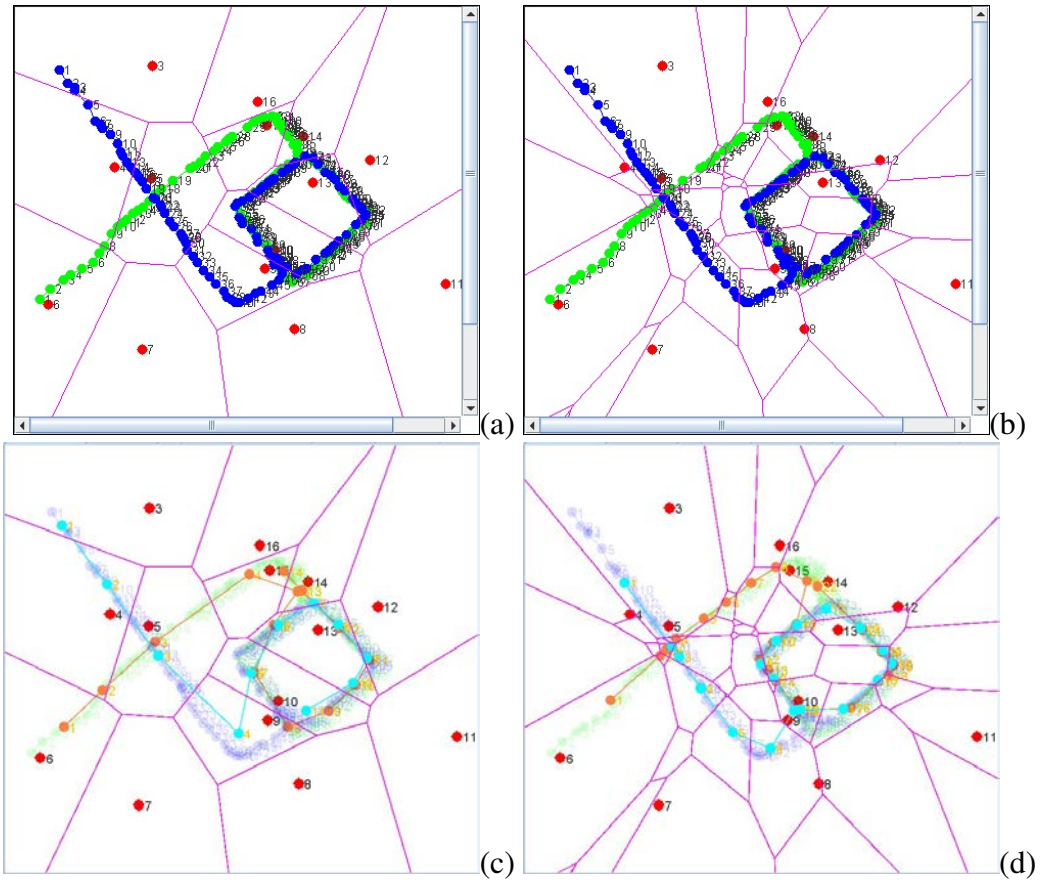


Fig. 7.15 OkVDs with two trajectories T_{31} , T_{32} and a set P_U of generators: (a) O1VD; (b) O5VD with two trajectory datasets; (c) O1VD with topologically simplified trajectories; (d) O5VD with topologically simplified trajectories.

Figure 7.15 displays HOVD of locations of the 14 clusters. Figure 7.15(a) shows O1VD with two tourist-group trajectory datasets while Figure 7.15(b) displays a corresponding O5VD. These clusters can be used to represent interesting POI P_U . These two trajectories' datasets in Figure 7.15 are recorded every 15 seconds for half an hour from two different

Emergency Management Application

tourist groups walking around the Cairns lagoon. The data collected from tourist T_{31} is represented in green, T_{32} is shown in blue in Figure 7.15. As P_U represents souvenir shops, people walking around the study region will have different nearest souvenir shops. Thus, when people pass through different paths, they will have a clear understanding of which area has the same HOI. Figure 7.15(c) and 7.15(d) show a series of OkVD with topologically simplified trajectories. T_{31} is represented in orange while T_{32} is in light-blue in Figure 7.15(c) and 7.15(d) respectively.

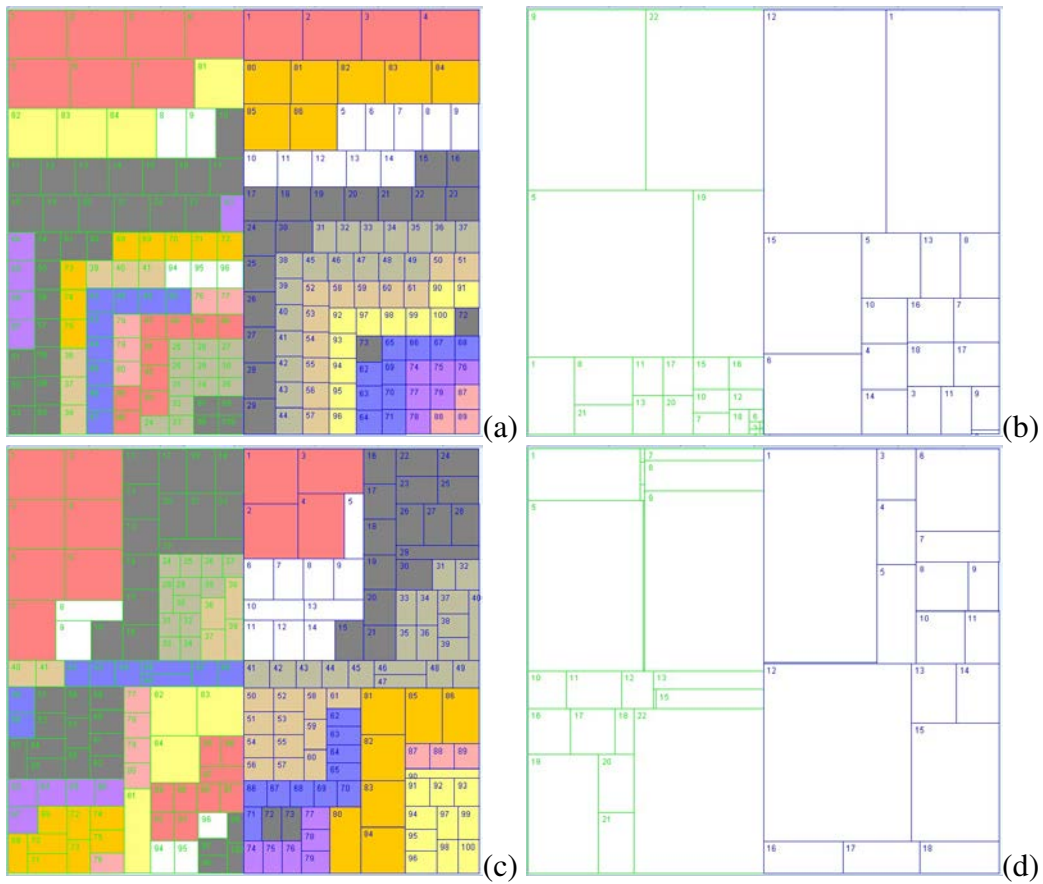


Fig. 7.16 Geometrical HOI and topologically simplified HOI treemaps with datasets shown in Figure 7.15: (a) squarified treemap with O1VD; (b) squarified treemap of topologically simplified datasets with O5VD; (c) pivot by split size treemap with O1VD; (d) pivot by split size treemap of topologically simplified datasets with O5VD.

Figure 7.16 shows some outcome results of treemaps based on the datasets shown in Figure 7.15. And Table 7.1 shows the aspect ratio of various treemap layouts and distance changes. In order to compare different treemap layout algorithms with these two metric

measures, we can measure the difference of these treemaps with T_{31} and T_{32} trajectories. In Table 7.1, we can clearly see that the squarified treemap has the smallest aspect ratio. Figure 7.16(a) and 7.16(b) display squarified treemaps in different k^{th} order information. Figure 7.16(b) can easily get clusters grouped on topologically simplified HOI, and the clustered squares are easier for people to use to compare which rectangle ($T_{31.9}$ and $T_{32.22}$) is more efficient and which is a more popular area for souvenir shops ($P_U(12, 13, 14, 15, 16)$). As we can also get this information from squarified treemap based on the original dataset, at the same time, we can also obtain the same information based on topologically simplified HOI treemaps. However, for a distance-change function, it is the worst one among these 7 treemaps, which means this treemap changes as k is changed for HOI.

As shown in Table 7.1, considering the best treemap design for various k^{th} order information when data is updated, pivot by split size treemap is the best one among these 7 treemaps. It has a low-aspect ratio and low-distance change. Figure 7.16(c) and 7.16(d) display O1VD with geometrical HOI and O5VD with topologically simplified HOI. Figure 7.16(d) shows more balanced layout information than we get from 7.16(b). But it is with a low distance change, which means this treemap layout gets lower rectangles movement as data is updated.

7.5 HODNA Application

We illustrate how the HODNAC and HODNAIF HOI can be used for emergency management using the dataset shown in Figure 7.7.

Figure 7.17 demonstrates HOVD and HODNAC. Figure 7.17(a) shows O1VD and 7.17(b) shows O5VD with three photo-taker trajectories. These three trajectory datasets in Figure 7.17(a) are collected every minute for half an hour from three different photo-taker groups who walked to the Cairns lagoon. The data collected from tourist T_{33} is represented in green, T_{34} is shown in blue, and T_{35} is depicted in light-pink in Figure 7.17(a). As P_U represents information centres, people moving around the study region will have different nearest information centres as they walk around. That is, people passing through different paths will have a good understanding of which area has the same HOI when they move along.

Table 7.1 Comparison with the Flickr example dataset.

Treemaps	Aspect Ratio		Average	Changes	Average
Binary	Geo	T_{31}	12.98	T_{31}	89.76
		T_{32}	9.03		
	Top	T_{31}	14.17	T_{32}	84.83
		T_{32}	10.05		
Ordered	Geo	T_{31}	2.17	T_{31}	121.09
		T_{32}	2.31		
	Top	T_{31}	4.37	T_{32}	116.57
		T_{32}	3.89		
PBS	Geo	T_{31}	2.21	T_{31}	151.72
		T_{32}	1.99		
	Top	T_{31}	3.44	T_{32}	163.56
		T_{32}	3.21		
PBSS	Geo	T_{31}	1.84	T_{31}	150.27
		T_{32}	1.83		
	Top	T_{31}	3.43	T_{32}	142.6
		T_{32}	3.76		
Slice	Geo	T_{31}	179.29	T_{31}	52.59
		T_{32}	200.39		
	Top	T_{31}	43.1	T_{32}	54.07
		T_{32}	40.48		
Squarified	Geo	T_{31}	1.2	T_{31}	203.45
		T_{32}	1.19		
	Top	T_{31}	1.5	T_{32}	201.82
		T_{32}	1.67		
Strip	Geo	T_{31}	1.73	T_{31}	135.87
		T_{32}	1.94		
	Top	T_{31}	3.27	T_{32}	152.46
		T_{32}	3.39		

Figure 7.17(c) shows HODNAC of trajectory T_{33} . As Figure 7.17(c) shows information that trajectory node $T_{33.14}$ and $T_{33.15}$ have the same TIF which is “1” and it is filled by a red rectangle. In another words, these two trajectory nodes get the same topological HOI. The highest OIF is when $k=13$. OIF is 0.765. The HOIF for T_{33} is 0.495. Figure 7.17(d) shows HODNAC of trajectory T_{34} . T_{34} does not show any two trajectory nodes with exactly the same topological HOI. As Figure 7.17(d) displays, the trajectory nodes $T_{34.16}$ and $T_{34.17}$ have

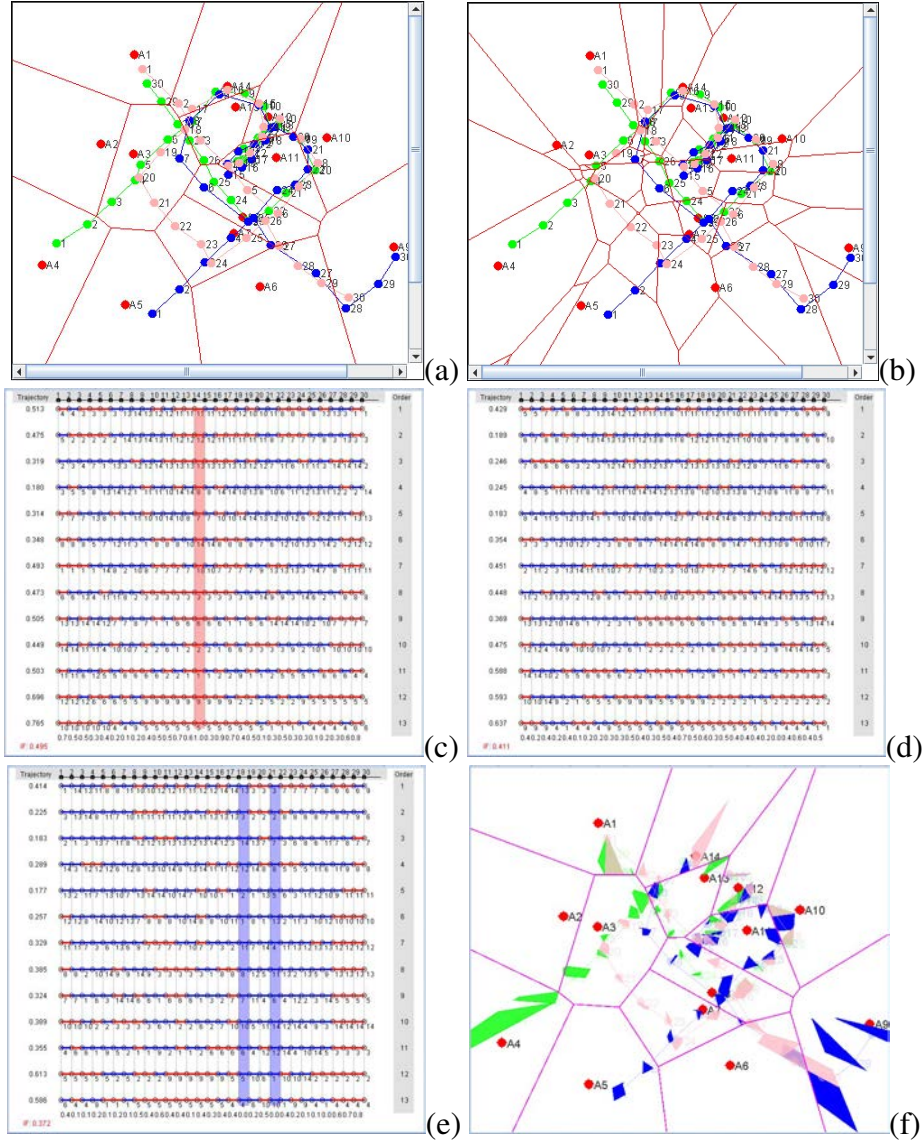


Fig. 7.17 HOVD and HODNAC: (a) O1VD; (b) O5VD with generator P_U and trajectory datasets T_{33} in green, T_{34} in blue and T_{35} in light-pink; (c) HODNAC for trajectory T_{33} ; (d) HODNAC for trajectory T_{34} ; (e) HODNAC for trajectory T_{35} ; (f) HOTA with trajectory area T_{33} in green, T_{34} in blue and T_{35} in light-pink.

the highest TIF, 0.7. The highest OIF is the same as T_{33} , when $k=13$, OIF is 0.637. The HOIF for T_{34} is lower than T_{33} : it is 0.411. Figure 7.17(e) draws HODNAC of trajectory T_{35} . Although T_{35} does not show any two trajectory nodes with completely the same topological HOI, T_{35} shows two TIF as “0”, which are $T_{35.18}$ and $T_{35.19}$; $T_{35.21}$ and $T_{35.22}$. In other words, these two sets of trajectory nodes have totally different topological HOIs, which means

people should be more focussed on the area where those trajectories located. The highest TIF is 0.8 for trajectory $T_{35.29}$ and $T_{35.30}$. The highest OIF is different compared to T_{33} and T_{34} , OIF for T_{35} is when $k=12$. Because trajectory T_{35} has very low TIF and OIF, the HOIF for T_{35} is 0.372.

We compare and contrast these three trajectory datasets T_{33} , T_{34} and T_{35} in Figure 7.17(c), (d) and (e). The area (around generator P_{U11}), $T_{33.14}$ and $T_{33.15}$, has a very high TIF relationship with the trajectory datasets. When trajectory nodes T_{34} and T_{35} are located in this area, the TIF of T_{34} and T_{35} is very high. In another words, trajectory datasets in this area have a strong relationship (Impact Factor) with POIs. On the other hand, T_{35} has two of the lowest TIF, which are $T_{35.18}$ and $T_{35.19}$; $T_{35.21}$ and $T_{35.22}$. The area (around generator P_{U3}) located within these trajectory nodes has the lower relationship (Impact Factor) POIs. For OIF, T_{33} the lowest is $k=4$, T_{34} the lowest is $k=5$, T_{35} the lowest is $k=5$. Based on the results, when $k \in (3, 4, 5)$, trajectory datasets have very low OIF. Overall, we can conclude about those three trajectory datasets, T_{33} has better relationships with POIs, when people make decisions, T_{33} is a better trajectory than T_{34} or T_{35} .

Figure 7.17(f) displays HOTA of three trajectories. T_{33} is represented in green, T_{34} is shown in blue, and T_{35} is depicted in light-pink. It shows the area that contains the same topological information of each trajectory node of NN information. It clearly shows that topological data based on POIs P_U are complex and tight around the Cairns CBD. Thus, depending on the research area and trajectories around the Cairns CBD, topological information contains very complex and sensitive relationships between POIs and trajectories.

7.6 Combination of HOPC and HOR for Emergency Management

We illustrate how a trajectory with geometrical, topological, and directional Mul-HOI can be used for emergency management using the dataset shown in Figure 7.2.

7.6.1 Mitigation

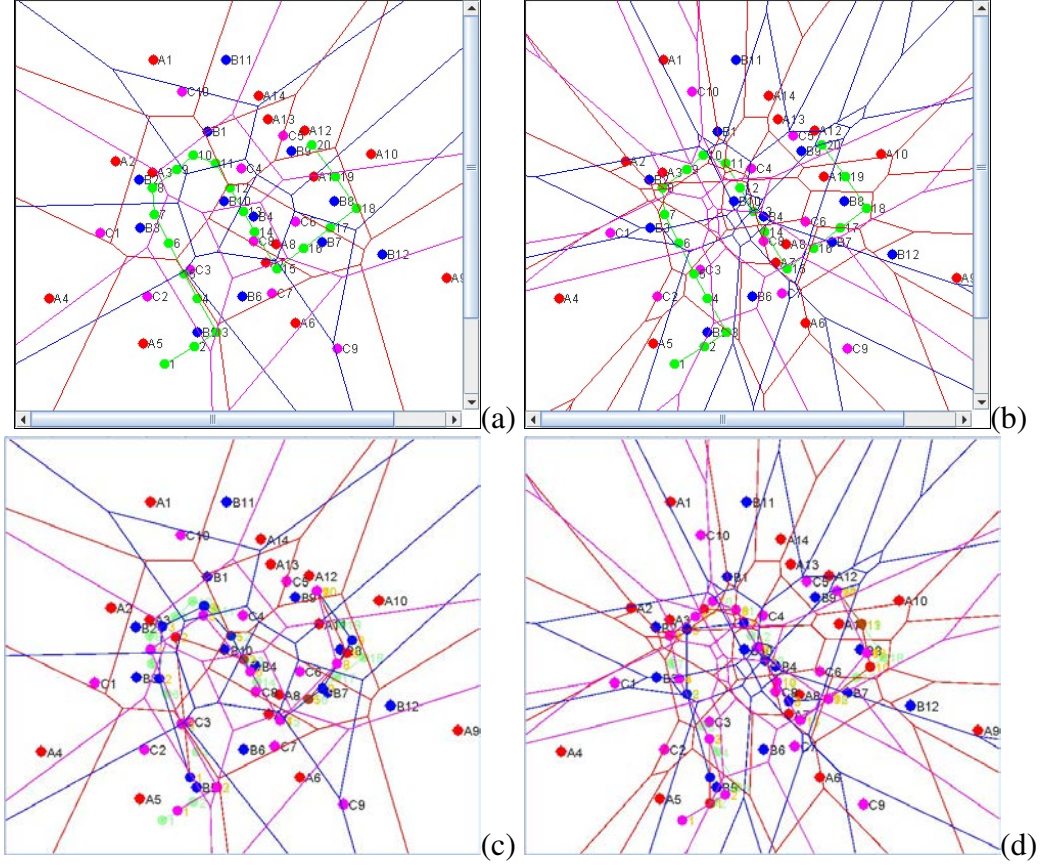


Fig. 7.18 HOVDs for trajectory T_{36} : (a) O1VD and (b) O5VD of trajectory T_{36} with Mul-HOI P_U, P_V, P_W ; (c) O1VD and (d) O5VD with 3 new topologically simplified trajectories. T_{36U} in red, T_{36V} in blue, whilst T_{36W} in magenta.

Figure 7.18(a) and Figure 7.18(b) show Order- k Voronoi diagrams with different values of k . A trajectory dataset is collected every minute for a period of 20 minutes from a tourist group who walked around the Cairns CBD. The data collected from the tourist group T_{36} is represented in green. Let us consider that P_U represents tourist attractions, P_V represents information centres, and P_W represents emergency centres. Figure 7.18(a) shows O1VD, and Figure 7.18(b) displays O5VD with only T_{36} data information. Considering the three different POIs datasets, k NNs for T_{36} are different for each POIs dataset. Mul-HOI can provide useful information to indicate where the higher-order k nearest help centres are along

the path of the trajectory, assuming that all three types of help centres include emergency centres, tourist attractions, and information centres.

Figure 7.18(c) and Figure 7.18(d) show different values of k with three topologically simplified trajectories. T_{36U} is represented in red, T_{36V} is shown in blue, and T_{36W} is in magenta. The local council can make decisions about the best k locations for setting up emergency tables or information tables for tourists before a disaster arrives. The k nearest neighbouring emergency units provide suggestions for areas where more activities occur. This provides useful information for tourists to make back-up plans when the nearest information centre is not available or functional. VA approaches provide a way to leverage data visualisation and human cognitive ability [117]. With the additional information from the topologically simplified trajectories, the local government can make more informed decisions for planning emergency supports before damage from the disaster occurs.

7.6.2 Preparedness

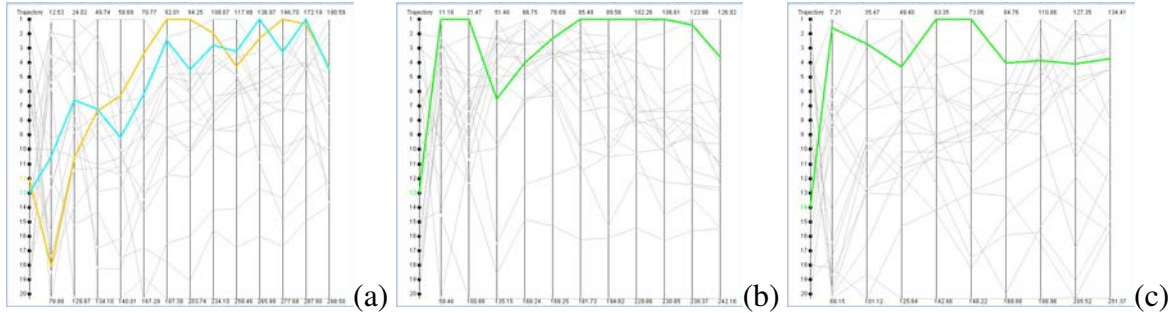


Fig. 7.19 Geometrical PCMMs: (a) Generator P_U 's geometrical PCMM for T_{36U} ; (b) Generator P_V 's geometrical PCMM for T_{36V} ; (c) Generator P_W 's geometrical PCMM for T_{36W} .

Figure 7.19, Figure 7.20, and Figure 7.21 show different values of k with Order- k information. Figure 7.19 displays geometrical information to determine the distance and relationship of the trajectory T_{36U} based on generators P_U , P_V and P_W . It also highlights the quantitative PCMM and qualitative PCMM information between Mul-HOI trajectory T_{36} and generators P_U , P_V and P_W . Figure 7.20 shows topological information to define the relationship of trajectory T_{36U} through generators P_U , P_V and P_W . It displays a relationship

7.6 Combination of HOPC and HOR for Emergency Management

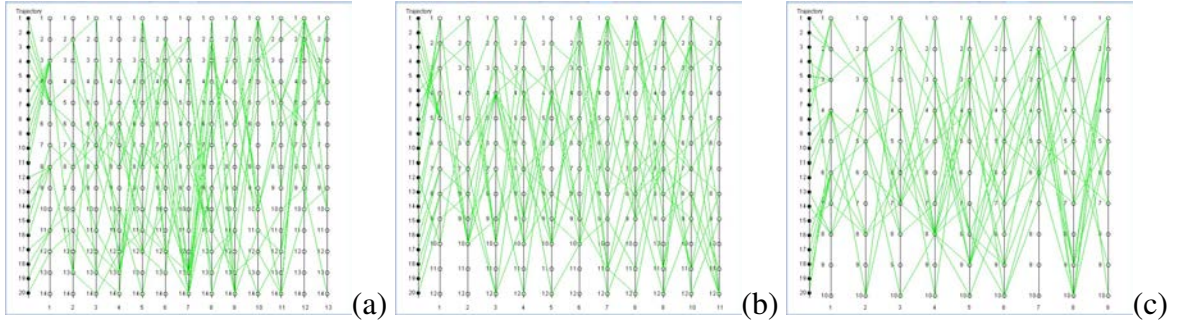


Fig. 7.20 Topological PCMMs: (a) Generator P_U 's topological PCMM for T_{36U} ; (b) Generator P_V 's topological PCMM for T_{36V} ; (c) Generator P_W 's topological PCMM for T_{36W} .

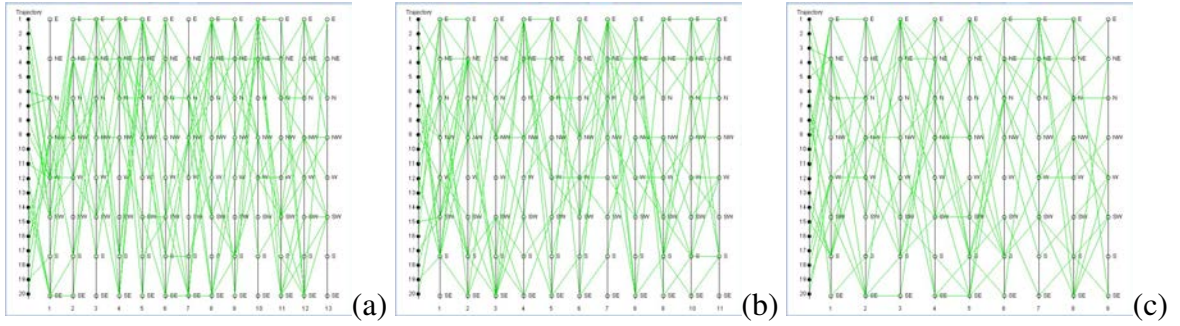


Fig. 7.21 Directional PCMMs: (a) Generator P_U 's directional PCMM for T_{36U} ; (b) Generator P_V 's directional PCMM for T_{36V} ; (c) Generator P_W 's directional PCMM for T_{36W} .

trend for different order k values between trajectory and generators. Figure 7.21 presents directional information to appraise the directional information of the trajectory T_{36U} based on generators P_U , P_V and P_W . It not only draws the trajectory directional information based on generators but also displays the most focused/popular direction within different order k .

Before a disaster or emergency situation occurs, PCMMs can help government units to focus on top-priority regions based on quantitative and qualitative PCMM where the emergency/information centre has the shortest distance to the trajectory. Secondly, government units also can focus on the most popular/sensitive generators based on topological PCMM where the emergency/information centre has the strongest relationship with the disaster trajectory. Moreover, they can also pay attention to the directions where directional PCMMs have strong relationships between trajectories and generators through different order k values. Thus, PCMM provides three different ways to analyse multiple sets of POIs to trajectory

Mul-HOI. It allows governments to suggest or provide better decisions on what (order), where (location), and which (centre) begins to provide the next emergency response.

7.6.3 Response

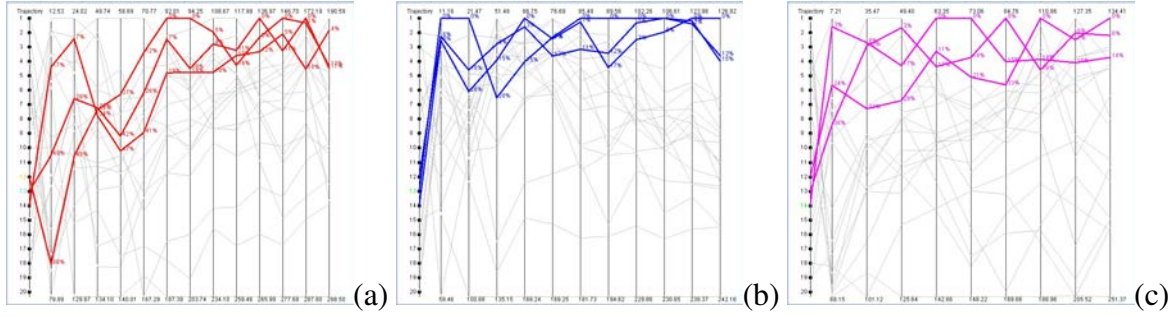


Fig. 7.22 Geometrical PCMMs: (a) Generator P_U 's geometrical PCMM for T_{36U} ; (b) Generator P_V 's geometrical PCMM for T_{36V} ; (c) Generator P_W 's geometrical PCMM for T_{36W} .

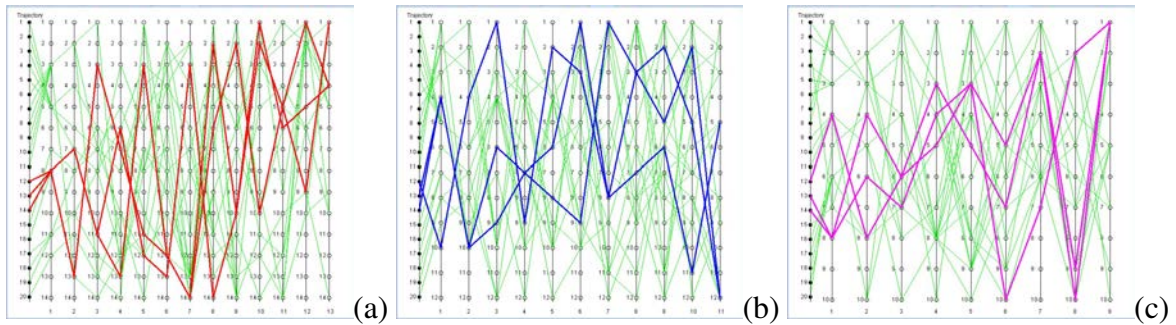


Fig. 7.23 Topological PCMMs: (a) Generator P_U 's topological PCMM for T_{36U} ; (b) generator P_V 's topological PCMM for T_{36V} ; (c) Generator P_W 's topological PCMM for T_{36W} .

Considering PCMM between trajectory T_{36U} and generators P_U , P_V , and P_W , geometrical PCMMs are shown in Figure 7.22, topological PCMMs are in Figure 7.23, and directional PCMMs in Figure 7.24. Clearly, we can see that although the numbers of clusters in the topologically simplified trajectories are different, corresponding HOI patterns are very similar. Figure 7.22(a) shows the common PCMM in red based on P_U . Figure 7.22(b) shows the common PCMM in blue based on P_V while Figure 7.22(c) shows the common PCMM

7.6 Combination of HOPC and HOR for Emergency Management

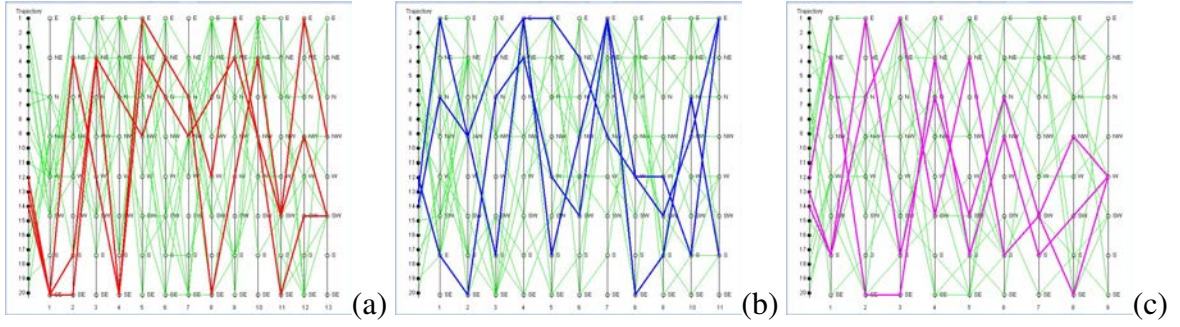


Fig. 7.24 Directional PCMMs: (a) Generator P_U 's directional PCMM for T_{36U} ; (b) Generator P_V 's directional PCMM for T_{36V} ; (c) Generator P_W 's directional PCMM for T_{36W} .

in magenta based on P_W . The quantitative and qualitative PCMMs with topological and directional information are shown in Figure 7.23 and Figure 7.24, respectively.

In the response to an emergency situation, better and faster solutions provided based on different issues are crucial. Response plans can be introduced based on patterns from PCMMs. From Figure 7.22, we can choose trajectory nodes $t_{12} - t_{14}$ as response plans since these nodes exhibit similar quantitative and qualitative geometrical PCMM. Interestingly, PCMMs with the geometrical information improve after $k=6$ as can be seen from Figure 7.22. Based on the topological information, generators ($P_{A3-(3,6,7,8,11,13)}$, $P_{B3-(1,4,6,7,9,10)}$, and $P_{C3-(3,4,5,6,7,8)}$) exhibit better topological PCMMs. One more thing to note is that the trajectory has the *East* side as a focusing direction for all datasets P_U , P_V , and P_W .

7.6.4 Recovery

Considering HORCs and HOk -Means between trajectory T_{36} and generators P_U , P_V , and P_W , with O1VD HRC information in Figure 7.25 and O5VD HRC information in Figure 7.26. Figure 7.25(a) displays O1VD HRC for P_U where HOI of T_{36U} is displayed in green. Figure 7.25(b) displays O1VD HRC for P_V where HOI of T_{36V} is presented in blue. Figure 7.25(c) displays O1VD HRC for P_W where HOI of T_{36W} is shown in magenta. Figure 7.25(d) displays O1VD HRC for P_U , P_V , and P_W where HOI of T_3 is displayed in red. Figure 7.26(a) displays O5VD HRC for P_U where HOI of T_{36U} is shown in green. Figure 7.26(b) displays O5VD HRC for P_V where HOI of T_{36V} is presented in blue. Figure 7.26(c)

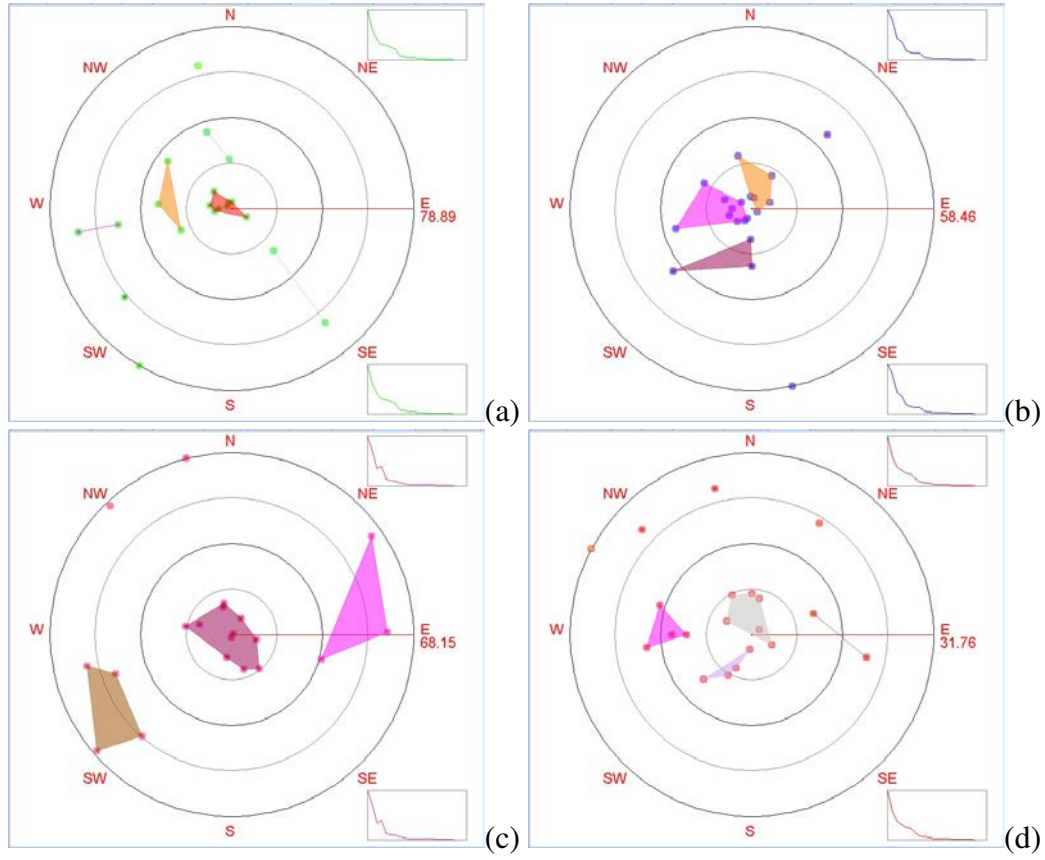


Fig. 7.25 HOCRs for O1VD: (a) Generator P_U 's HOCR for T_{36U} ; (b) Generator P_V 's HOCR for T_{36V} ; (c) Generator P_W 's HOCR for T_{36W} ; (d) Generator P_U , P_V , and P_W 's HOCR for T_{36} .

displays O5VD HOCR for P_W where HOI of T_{36W} is drawn in magenta. Figure 7.26(d) displays O5VD HOCR for P_U , P_V , and P_W where HOI of T_3 is displayed in red.

We can see that HOCRs show the POIs are mostly located on the *West* side and not close to each other based on P_U 's O1VD information shown in Figure 7.25(a). In Figure 7.25(b), a HOCR shows that the POIs are mostly located on the *West* side and very close to the trajectory based on P_V 's O1VD information. In Figure 7.25(c), a HOCR displays that the POIs remain in *East* and *West* directions and a big cluster is very close to the trajectory (less than 25% of *Max* distance) based on P_W 's O1VD information. In Figure 7.25(d), a HOCR shows that the POIs are mostly located on the *West* side and very close to the trajectory based on all P_U , P_V , and P_W 's O1VD information. However, considering HOCR for O5VD information, the POIs P_U are mostly located on the *East* side and very close to the trajectory

7.6 Combination of HOPC and HOR for Emergency Management

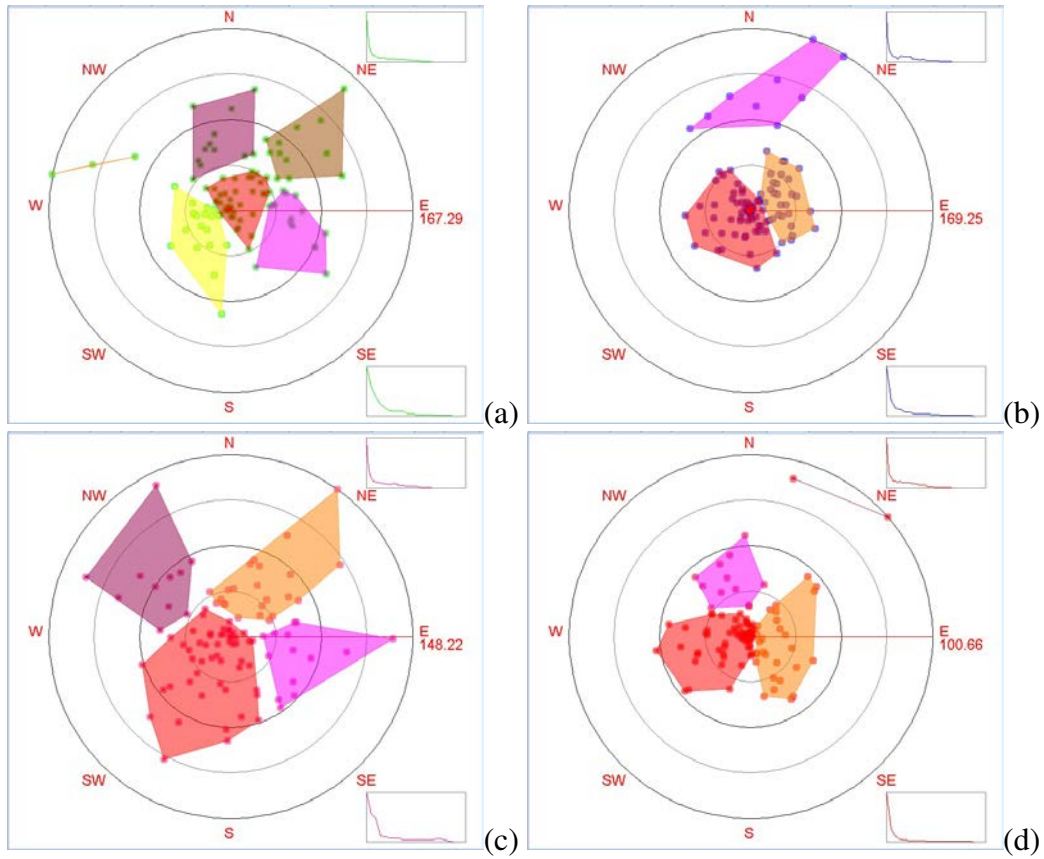


Fig. 7.26 HOPCs for O5VD: (a) Generator P_U 's HOPC for T_{36U} ; (b) Generator P_V 's HOPC for T_{36V} ; (c) Generator P_W 's HOPC for T_{36W} ; (d) Generator P_U , P_V , and P_W 's HOPC for T_{36} .

(less than 50% of Max distance) as shown in Figure 7.26(a). In Figure 7.26(b), the POIs P_V are very close to the trajectory (less than 25% of Max distance). In Figure 7.26(c), the POIs P_W are averagely located around the trajectory. In Figure 7.26(d), HOPC shows that the POIs are closely located around the trajectory based on all P_U , P_V , and P_W 's O5VD information.

Clustering results for order-1 information are displayed in Figure 7.27(a) with the original elbow method, and in Figure 7.27(b) with the HO1-Means method. As we are only looking for the 1st nearest neighbour HOI, the result will be the same because it calculates only the first nearest neighbour information. In Figure 7.27, we can see that in the elbow chart showing the SSE after running k -means, clustering for k going from 1 to 19 is evident. We see a clear elbow P_U very similar to elbow ALL . Elbow P_V is lower than P_U and ALL , and Elbow

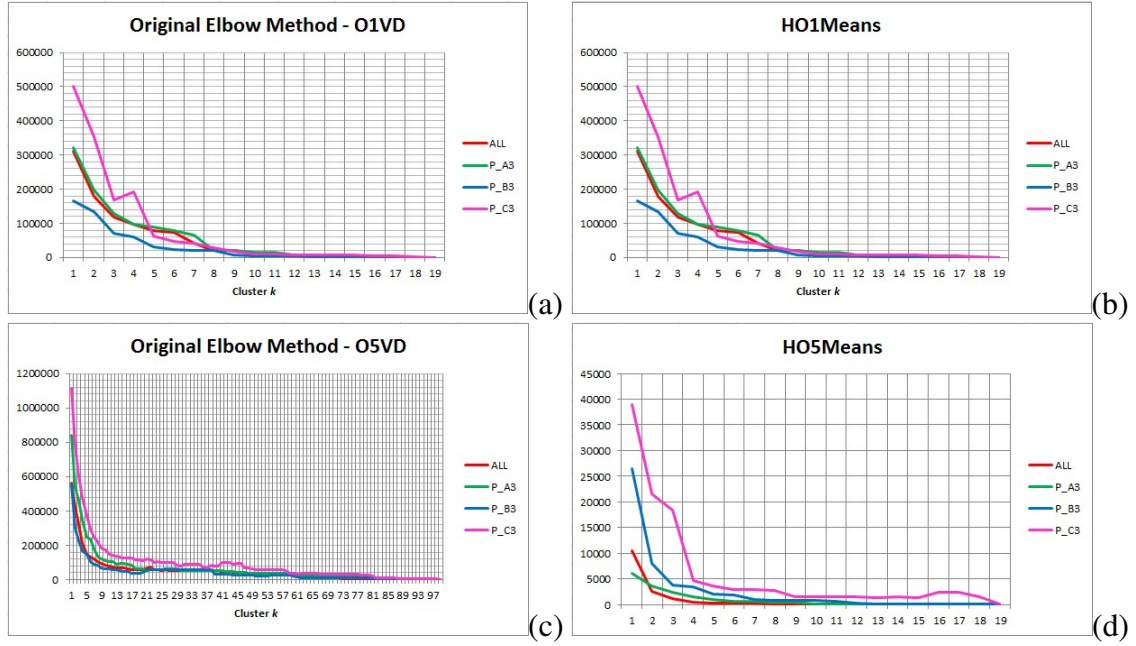


Fig. 7.27 Distributions of SSE for HOI (x-axis: order k ; y-axis: SSE) (a) Original elbow method for O1VD; (b) HO1-Means for O1VD; (c) Original elbow method for O5VD; (d) HO5-Means of trajectory dataset T_3 with generators P_U , P_V and P_W .

P_W is higher than P_U and ALL . Thus, we can combine P_U and ALL together to undertake future analysis.

When we consider HOI for $k=5$, Figure 7.27(c) shows the original elbow method while Figure 7.27(d) shows HO5-Means. Note how the elbow chart for the original method does not have a clear elbow. It is a fairly smooth curve, and it is unclear what is the best value for the number of clusters k . This result shows that the original method of k -means clustering does not work very well for trajectory Mul-HOI. On the other hand, Figure 7.27(d) presents the results by HO5-Means, it suggests that the optimal numbers of clusters of P_U , P_V , P_W and all together that the HO5-Means are $k=6$ for P_U , $k=3$ for P_V , $k=4$ for P_W , and $k=4$ for all generators together. Interestingly, the elbow P_U is very similar to the elbow ALL in HO5-Means although P_V and P_W are higher. Therefore, the proposed HO_k -means method is superior to the original method for selecting the best k for clustering of trajectory HOI.

In short-term emergency recovery, we can choose different numbers of clusters as recovery plans based on HORC. For the long term, recovery plans can be introduced based on patterns

from PCMM and HOK-Means, the k^{th} of Mul-HOI and the best k for clustering trajectory HOI. Therefore, the best choice for order- k and cluster number k is a solid reference for decision-making and is useful for the recovery phase of emergency management.

7.7 Conclusion

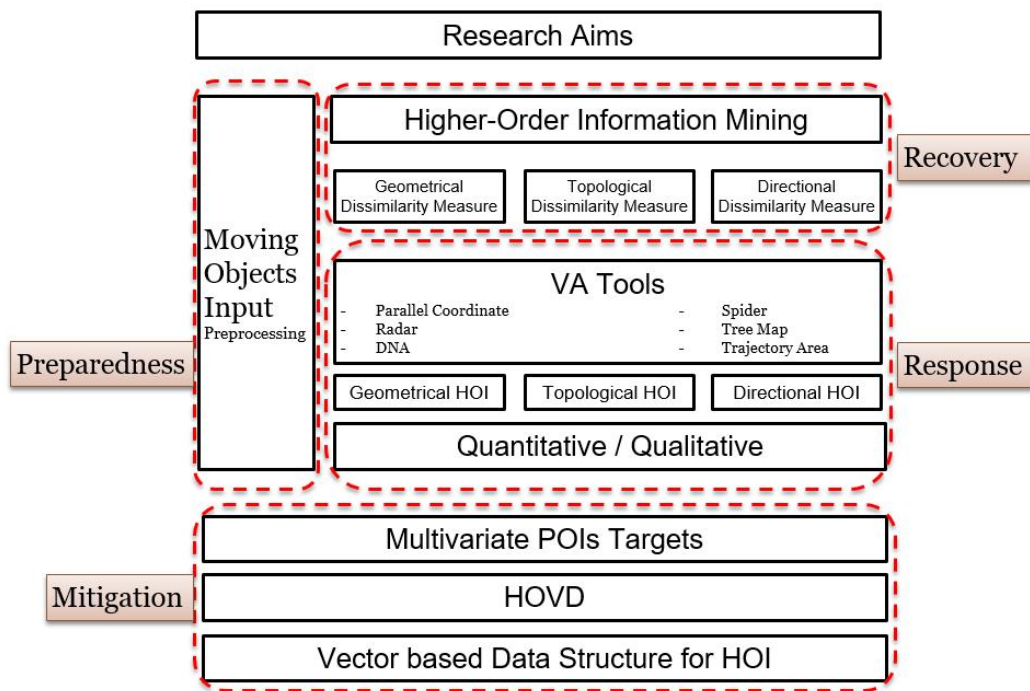


Fig. 7.28 Emergency management research framework.

In conclusion of this chapter, we propose new visual analytical approaches to analyse trajectories with multivariate datasets for emergency management. The unified data structure gives general suggestions of “what-if” information before making plans in all stages of emergency management. Figure 7.28 depicts the details of the proposed framework for this research in Chapter 3. The proposed research frameworks and approaches are applied to the emergency management. Mitigation phase is a realisation that eliminates or decreases the causes of disasters or risks. The proposed framework provides vector based data structure HOI, which not only focuses on HOVD, but also supports multivariate POIs targets. In the preparedness phase, it builds the supporting environment and prepares plans to save

Emergency Management Application

lives and minimise disaster damage such as having emergency personnel on standby. The proposed framework provides the data input preprocessing methodologies, which can prepare and pre-build the datasets to be ready to use for further analysis. Response phase consists of providing activities following the emergency situation to assist disaster victims. The proposed framework provides 5 different VA tools which can focus on the different scenarios. Depending on the user or expert's requirements, different approaches can provide geometrical, topological and directional HOI of quantitative and qualitative information. Recovery happens when the disaster is over. It is a set of activities that returns all systems back to normal and also prevents follow up damage. The proposed framework provides 3 different types of dissimilarity measurements and HOI data mining techniques, which returns top- k dissimilarities and clusters of HOI.

Chapter 8

Conclusion

In this chapter we conclude our work, highlight the main contributions that this thesis makes, and then present some possible future avenues for research.

8.1 Background

As our environment has been changing into data-rich and information-rich which is a revelatory exploration of the hottest trend in technology and the dramatic impact it will have on the economy, science and society at large. It intends to explain where we are, traces how we got here, and offers an urgent need to the benefits and dangers that lie ahead. Large datasets refers to people's new found ability to crunch a vast quantity of information, to analyse it instantly and draw sometimes astonishing conclusions from it. Currently, due to advances in technologies and the explosion of rich data, the uncertainty of information in decision-making in the modern society has become a significant issue. What is the best way to make different decisions based on different situations, will stay with our real-world all the time. "What-if" analysis is the study of how the output of an analysis model varies with changes of the inputs. It is particularly useful for decision-making when uncertainties are present.

Disasters and emergencies can lead to various forms of financial, structural, and environmental damage. Even though it is almost impossible to avoid occurrences of disasters,

Conclusion

effective prediction and preparedness along with a post-emergency management program can mitigate the risk and damage. Trajectory analysis has emerged in the last decade as a very active research domain with many potential applications for emergency management. HOI of moving objects is of great importance for processing “what-if” scenarios in emergency management. Mul-HOI management is a crucial key to the success of emergency management since emergency management involves developing plans with a given set of multiple resources (multivariate).

Due to the complexity of trajectory datasets and related data analysis, there is a need to develop new and innovative ways to allow researchers and data analysts to reveal hidden information and knowledge embedded in trajectory datasets. While most of the existing research efforts are focused on univariate data only, it cannot support trajectory datasets, it is non-interactivity, it only focuses on quantitative information, there is no higher-order dissimilarity measurement method, and is without higher-order or other data mining techniques provided. In terms of different types of HOI, there are some approaches for geometrical HOI, but there is very limited research on topological and directional HOI.

The first contribution of this thesis is to identify and define unified data structure that supports Mul-HOI for POIs and moving objects. The vector-based data structure allows datasets between POIs and trajectories to be modelled and quantified. The second contribution is to identify and define two ways of data preprocessing of trajectory, which are topologically simplified and topologically interpolated. The data preprocessing could allow people to choose the best method based on given trajectory dataset attributes. The third contribution of this thesis is to introduce five different VA approaches in which HOPC, HOS, HOR, HOTrM and HODNA are quantified and qualified Mul-HOI with regard to three categories of k NN and k OR information; these approaches enable users to interactively analyse a large volume of trajectories with different POIs, and to understand the HOI relationship of POIs (emergency resources) and trajectories (disasters). The fourth contribution is to develop dissimilarity measurement approaches and HOk -Means clustering methods for analysing large trajectory dataset and HOI in particular for emergency management. This research is efficient and effective to provide a new way to interactively visualise Mul-HOI for moving

objects, and to provide semantic answers to problems such as top- k trajectory HOI and “what-if” analysis for emergency management.

8.2 Summary of Contributions

The goal of this research is to provide a unified data structure for multivariate support of trajectory data analysis and investigate the application of VA tools to provide various HOI with multivariate trajectory datasets of POIs for moving objects. We propose a set of VA tools for exploring and analysing Mul-HOI with datasets of moving objects.

In summary, the main contributions of this research are as follows:

- provision of unified vector-based data structure for supporting multivariate POI targets, and for efficiently manipulating datasets with dynamically moving objects;
- introduction of input data preprocessing methods for supporting Mul-HOI;
- proposal of new VA tools based on parallel coordinates, spider charts, radar charts, treemaps, and DNA charts for analysing large trajectory datasets with three categories (geometrical, topological and directional) of Mul-HOI;
- design and implementation of a new VA tool called HOPC for the determination of PCMM to quantify both qualitative and quantitative Mul-HOI with regard to k NN and k OR information, to provide all geometrical, topological and directional HOI, and to introduce geometrical dissimilarity measure of Mul-HOI;
- design and implementation of a new VA tool called HOS for the determination of DBR and DDM to quantify both qualitative and quantitative Mul-HOI with regards to k NN and k OR information, to innovate new directional HOI, and to introduce directional dissimilarity measures of Mul-HOI;
- design and implementation of a new VA tool called HOR for the determination of HO k -Means clustering for trajectory data mining to quantify both qualitative and

Conclusion

quantitative Mul-HOI with regard to k NN and k OR information, and to provide all geometrical, topological and directional HOI;

- design and implementation of a new VA tool called HOTrM for the determination of seven different types of treemap methods to quantify the best presentation method for both qualitative and quantitative Mul-HOI with regard to k NN and k OR information;
- design and implementation of a new VA tool called HODNA for the identification of a moving object's unique topological colour-code and the determination of HOTA and HODNAIF to quantify both qualitative and quantitative Mul-HOI with regard to k NN and k OR information, and to identify and compare trajectory datasets based on topological impact factor;
- provision of a case study to demonstrate the usefulness and applicability of these new VA tools for emergency management and research framework to be able to improve and support emergency management of four basic phases.

8.3 Future Work

The outcomes of this thesis represent the first exploration of new insights and techniques, incorporating studies and approaches that break new ground. Accordingly, there are a number of limitations that can be addressed by future extensions to this research. The unified data structure for multivariate datasets that can support Mul-HOI for POIs and moving objects in Chapter 4 has only been validated as part of the research in this thesis. Future research can compare and contrast the results between topologically simplified and topologically interpolated method. DNA method can be extended to calculate the unique DNA colour-code to be able to locate trajectory datasets, and it can be used in information security transformation areas etc. Another possible direction can extend this thesis by implementing 3D VA tools for Mul-HOI and the proposed VA data structure that supports a wider range of interface for people to continue exploration with ease.

The research presented in this thesis could be applicable to a wide range of other applications that may lead to economic, social and environmental advantages. We have presented a case study using our techniques for emergency management and tourism analysis. However, the techniques could also be applicable to geospatial datasets in a variety of other fields including crime analysis, epidemiology, business intelligence, geology, environmental monitoring, marketing and e-commerce. We wish to undertake further case studies to investigate the potential use of our framework in these other disciplines.

References

- [1] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. Visual analytics: Definition, process, and challenges. In *Information visualization*, pages 154–175. Springer, 2008.
- [2] Panos Kalnis, Nikos Mamoulis, and Spiridon Bakiras. On discovering moving clusters in spatio-temporal data. In *International Symposium on Spatial and Temporal Databases*, pages 364–381. Springer, 2005.
- [3] Ickjai Lee and Kyungmi Lee. A generic triangle-based data structure of the complete set of higher order voronoi diagrams for emergency management. *Computers, Environment and Urban Systems*, 33(2):90–99, 2009.
- [4] Ye Wang, Kyungmi Lee, and Ickjai Lee. Representing ordered order-k voronoi diagrams. In *Advanced Information Management and Service (ICIPM), 2011 7th International Conference on*, pages 71–75. IEEE, 2011.
- [5] 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, 2008.
- [6] Gennady Andrienko, Natalia Andrienko, Christophe Hurter, Salvatore Rinzivillo, and Stefan Wrobel. From movement tracks through events to places: Extracting and characterizing significant places from mobility data. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 161–170. IEEE, 2011.
- [7] Gerhard Albers and Thomas Roos. Voronoi diagrams of moving points in higher dimensional spaces. *Algorithm Theory—SWAT’92*, pages 399–409, 1992.
- [8] Leonidas Guibas, Joseph Mitchell, and Thomas Roos. Voronoi diagrams of moving points in the plane. In *Graph-Theoretic Concepts in Computer Science*, pages 113–125. Springer, 1992.
- [9] Gennady Andrienko, Natalia Andrienko, Urska Demsar, Doris Dransch, Jason Dykes, Sara Irina Fabrikant, Mikael Jern, Menno-Jan Kraak, Heidrun Schumann, and Christian Tominski. Space, time and visual analytics. *International Journal of Geographical Information Science*, 24(10):1577–1600, 2010.
- [10] Gennady Andrienko, Natalia Andrienko, and Stefan Wrobel. Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, 9(2):38–46, 2007.

References

- [11] James Dean Palmer. Using line and texture to visualize higher-order voronoi diagrams. In *Voronoi Diagrams in Science and Engineering, 2006. ISVD'06. 3rd International Symposium on*, pages 166–172. IEEE, 2006.
- [12] Alexandru Telea and Jarke J Van Wijk. Visualization of generalized voronoi diagrams. *Proc. IEEE VisSym*, 1:165–174, 2001.
- [13] Philip Ball. Data visualization: picture this. *Nature*, 418(6893):11–13, 2002.
- [14] Yu Zheng. Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3):29, 2015.
- [15] Liang Lin, Yongyi Lu, Yan Pan, and Xiaowu Chen. Integrating graph partitioning and matching for trajectory analysis in video surveillance. *IEEE transactions on Image Processing*, 21(12):4844–4857, 2012.
- [16] Brendan Tran Morris and Mohan Manubhai Trivedi. A survey of vision-based trajectory learning and analysis for surveillance. *IEEE transactions on circuits and systems for video technology*, 18(8):1114–1127, 2008.
- [17] Rebecca N Handcock, Dave L Swain, Greg J Bishop-Hurley, Kym P Patison, Tim Wark, Philip Valencia, Peter Corke, and Christopher J O'Neill. Monitoring animal behaviour and environmental interactions using wireless sensor networks, gps collars and satellite remote sensing. *Sensors*, 9(5):3586–3603, 2009.
- [18] 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 *Visualization Symposium (PacificVis), 2014 IEEE Pacific*, pages 137–144. IEEE, 2014.
- [19] 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, 2013.
- [20] Gopal Pingali, Agata Opalach, Yves Jean, and Ingrid Carlbom. Visualization of sports using motion trajectories: providing insights into performance, style, and strategy. In *Proceedings of the conference on Visualization'01*, pages 75–82. IEEE Computer Society, 2001.
- [21] Mirela C Popa, Léon JM Rothkrantz, Caifeng Shan, Tommaso Gritti, and Pascal Wiggers. Semantic assessment of shopping behavior using trajectories, shopping related actions, and context information. *Pattern Recognition Letters*, 34(7):809–819, 2013.
- [22] Ickjai Lee, Reece Pershouse, Peter Phillips, and Chris Christensen. What-if emergency management system: a generalized voronoi diagram approach. In *Pacific-Asia Workshop on Intelligence and Security Informatics*, pages 58–69. Springer, 2007.
- [23] Yu Zheng and Xiaofang Zhou. *Computing with spatial trajectories*. Springer Science & Business Media, 2011.

-
- [24] Jörg Sander, Martin Ester, Hans-Peter Kriegel, and Xiaowei Xu. Density-based clustering in spatial databases: The algorithm gbscan and its applications. *Data mining and knowledge discovery*, 2(2):169–194, 1998.
 - [25] Danzhou Liu, Ee-Peng Lim, and Wee-Keong Ng. Efficient k nearest neighbor queries on remote spatial databases using range estimation. In *Scientific and Statistical Database Management, 2002. Proceedings. 14th International Conference on*, pages 121–130. IEEE, 2002.
 - [26] Ralf Hartmut Güting, Thomas Behr, and Jianqiu Xu. Efficient k-nearest neighbor search on moving object trajectories. *The VLDB Journal—The International Journal on Very Large Data Bases*, 19(5):687–714, 2010.
 - [27] Guochen Cai, Chihiro Hio, Luke Bermingham, Kyungmi Lee, and Ickjai Lee. Sequential pattern mining of geo-tagged photos with an arbitrary regions-of-interest detection method. *Expert Systems with Applications*, 41(7):3514–3526, 2014.
 - [28] Leila Hashemi Beni, Mir Abolfazl Mostafavi, Jacynthe Pouliot, and Marina Gavrilova. Toward 3d spatial dynamic field simulation within gis using kinetic voronoi diagram and delaunay tetrahedralization. *International Journal of Geographical Information Science*, 25(1):25–50, 2011.
 - [29] Xiaopeng Xiong, Mohamed F Mokbel, and Walid G Aref. Sea-cnn: Scalable processing of continuous k-nearest neighbor queries in spatio-temporal databases. In *Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on*, pages 643–654. IEEE, 2005.
 - [30] LS Jakobson and Melvyn A Goodale. Factors affecting higher-order movement planning: a kinematic analysis of human prehension. *Experimental brain research*, 86(1):199–208, 1991.
 - [31] Fosca Giannotti, Mirco Nanni, Fabio Pinelli, and Dino Pedreschi. Trajectory pattern mining. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 330–339. ACM, 2007.
 - [32] Pavel Berkhin. A survey of clustering data mining techniques. In *Grouping multidimensional data*, pages 25–71. Springer, 2006.
 - [33] Ben Shneiderman. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on graphics (TOG)*, 11(1):92–99, 1992.
 - [34] Brian Johnson. Treemap visualization of hierarchically structured information. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 369–370. ACM, 1992.
 - [35] Brian Johnson and Ben Shneiderman. Tree-maps: A space-filling approach to the visualization of hierarchical information structures. In *Proceedings of the 2nd conference on Visualization'91*, pages 284–291. IEEE Computer Society Press, 1991.
 - [36] Martin Wattenberg. Visualizing the stock market. In *CHI'99 extended abstracts on Human factors in computing systems*, pages 188–189. ACM, 1999.

References

- [37] Toshiyuki Asahi, David Turo, and Ben Shneiderman. Using treemaps to visualize the analytic hierarchy process. *Information Systems Research*, 6(4):357–375, 1995.
- [38] Andrew Fiore and Marc A Smith. Treemap visualizations of newsgroups. *Technical Report, Microsoft Research, Microsoft Corporation: Redmond, WA*, 2001.
- [39] Shixian Chu, Jinfeng Chen, Zonghuan Wu, Chee-Hung Chu, and Vijay Raghavan. A treemap-based result interface for search engine users. *Human Interface and the Management of Information. Methods, Techniques and Tools in Information Design*, pages 401–410, 2007.
- [40] Roberto Pinho, MC Ferreira de Oliveira, Rosane Minghim, and Marinho Gomes Andrade. Voromap: A voronoi-based tool for visual exploration of multi-dimensional data. In *Information Visualization, 2006. IV 2006. Tenth International Conference on*, pages 39–44. IEEE, 2006.
- [41] Stanley Smith Stevens. On the theory of scales of measurement. 1946.
- [42] Donna M Mertens. *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. Sage publications, 2014.
- [43] Shashi Shekhar, Sanjay Chawla, Sivakumar Ravada, Andrew Fetterer, Xuan Liu, and Chang-tien Lu. Spatial databases-accomplishments and research needs. *IEEE transactions on knowledge and data engineering*, 11(1):45–55, 1999.
- [44] National Research Council et al. *Improving disaster management: the role of IT in mitigation, preparedness, response, and recovery*. National Academies Press, 2007.
- [45] Thao P Nghiem, Agustinus Borgy Waluyo, and David Taniar. A pure peer-to-peer approach for knn query processing in mobile ad hoc networks. *Personal and Ubiquitous Computing*, 17(5):973–985, 2013.
- [46] Sarana Nutanong, Mohammed Eunus Ali, Egemen Tanin, and Kyriakos Mouratidis. Dynamic nearest neighbor queries in euclidean space. 2015.
- [47] François Anton, Darka Mioc, and Christopher Gold. The voronoi diagram of circles and its application to the visualization of the growth of particles. *Transactions on Computational Science III*, pages 20–54, 2009.
- [48] Sunil Arya, David M Mount, Nathan S Netanyahu, Ruth Silverman, and Angela Y Wu. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM (JACM)*, 45(6):891–923, 1998.
- [49] Jean-Daniel Boissonnat, Camille Wormser, and Mariette Yvinec. Curved voronoi diagrams. In *Effective Computational Geometry for Curves and Surfaces*, pages 67–116. Springer, 2006.
- [50] Gísli R Hjaltason and Hanan Samet. Distance browsing in spatial databases. *ACM Transactions on Database Systems (TODS)*, 24(2):265–318, 1999.
- [51] Christian A Lang and Ambuj K Singh. A framework for accelerating high-dimensional nn-queries. Technical report, Technical Report TRCS01-04, University of California, Santa Barbara, 2001.

-
- [52] Niranjana Mayya and VT Rajan. Voronoi diagrams of polygons: A framework for shape representation. *Journal of Mathematical Imaging and Vision*, 6(4):355–378, 1996.
- [53] Nick Roussopoulos, Stephen Kelley, and Frédéric Vincent. Nearest neighbor queries. In *ACM sigmod record*, volume 24, pages 71–79. ACM, 1995.
- [54] Cui Yu, Beng Chin Ooi, Kian-Lee Tan, and HV Jagadish. Indexing the distance: An efficient method to knn processing. In *Vldb*, volume 1, pages 421–430, 2001.
- [55] Stefan Berchtold, Daniel A Keim, H-P Kriegel, and Thomas Seidl. Indexing the solution space: A new technique for nearest neighbor search in high-dimensional space. *IEEE Transactions on Knowledge and Data Engineering*, 12(1):45–57, 2000.
- [56] Maurício Carvalho Martins de Paulo, Antônio Miguel Vieira Monteiro, and Eduardo Gerbi Camargo. Open source implementation of the multiplicatively weighted voronoi diagram as a terraview plugin. In *GeoInfo*, pages 39–48, 2011.
- [57] Tao Huang and MR Kamal. Morphological modeling of polymer solidification. *Polymer Engineering & Science*, 40(8):1796–1808, 2000.
- [58] Carlos RV Kiffer, Eduardo CG Camargo, Silvia E Shimakura, Paulo J Ribeiro, Trevor C Bailey, Antonio CC Pignatari, and Antonio MV Monteiro. A spatial approach for the epidemiology of antibiotic use and resistance in community-based studies: the emergence of urban clusters of escherichia coli quinolone resistance in sao paulo, brasil. *International journal of health geographics*, 10(1):17, 2011.
- [59] Amirtham Rajagopal, Paul Fischer, Ellen Kuhl, and Paul Steinmann. Natural element analysis of the cahn–hilliard phase-field model. *Computational mechanics*, 46(3):471–493, 2010.
- [60] Avneesh Sud, Erik Andersen, Sean Curtis, Ming C Lin, and Dinesh Manocha. Real-time path planning in dynamic virtual environments using multiagent navigation graphs. *IEEE transactions on visualization and computer graphics*, 14(3):526–538, 2008.
- [61] Ickjai Lee, Christopher Torpelund-Bruin, and Kyungmi Lee. Map segmentation for geospatial data mining through generalized higher-order voronoi diagrams with sequential scan algorithms. *Expert Systems with Applications*, 39(12):11135–11148, 2012.
- [62] Atsuyuki Okabe, Barry Boots, Kokichi Sugihara, and Sung Nok Chiu. *Spatial tessellations: concepts and applications of Voronoi diagrams*, volume 501. John Wiley & Sons, 2009.
- [63] Graeme John Hugo, Peter John Smailes, Colin Macgregor, Mark Fenton, and David Brunckhorst. Defining social catchments in non-metropolitan australia. 2001.
- [64] Miao Wang and Zhongxiao Hao. Nearest neighbors and continuous nearest neighbor queries based on voronoi diagrams. *Information Technology Journal*, 9(7):1467–1475, 2010.

References

- [65] Nicholas S Kalson, Tom Jenks, Maralyn Woodford, Fiona E Lecky, and Ken W Dunn. Burns represent a significant proportion of the total serious trauma workload in england and wales. *Burns*, 38(3):330–339, 2012.
- [66] Christoph Kleinn and Franti ek Vilčko. Design-unbiased estimation for point-to-tree distance sampling. *Canadian Journal of Forest Research*, 36(6):1407–1414, 2006.
- [67] Der-Tsai Lee. On k-nearest neighbor voronoi diagrams in the plane. *IEEE Transactions on Computers*, 100(6):478–487, 1982.
- [68] Mark Gahegan and Ickjai Lee. Data structures and algorithms to support interactive spatial analysis using dynamic voronoi diagrams. *Computers, environment and urban systems*, 24(6):509–537, 2000.
- [69] Ickjai Lee and Christopher Torpelund-Bruin. Geographic knowledge discovery from web map segmentation through generalized voronoi diagrams. *Expert Systems with Applications*, 39(10):9376–9388, 2012.
- [70] Martin Erwig, Ralf Hartmut Gu, Markus Schneider, Michalis Vazirgiannis, et al. Spatio-temporal data types: An approach to modeling and querying moving objects in databases. *GeoInformatica*, 3(3):269–296, 1999.
- [71] Charles Reiss, Alexey Tumanov, Gregory R Ganger, Randy H Katz, and Michael A Kozuch. Heterogeneity and dynamicity of clouds at scale: Google trace analysis. In *Proceedings of the Third ACM Symposium on Cloud Computing*, page 7. ACM, 2012.
- [72] Wolfgang Aigner, Silvia Miksch, Wolfgang Müller, Heidrun Schumann, and Christian Tominski. Visualizing time-oriented data—a systematic view. *Computers & Graphics*, 31(3):401–409, 2007.
- [73] Doreen Massey. *Space, place and gender*. John Wiley & Sons, 2013.
- [74] John Jobson. *Applied multivariate data analysis: volume II: Categorical and Multivariate Methods*. Springer Science & Business Media, 2012.
- [75] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE, 1996.
- [76] Jörn Kohlhammer, Daniel Keim, Margit Pohl, Giuseppe Santucci, and Gennady Andrienko. Solving problems with visual analytics. *Procedia Computer Science*, 7:117–120, 2011.
- [77] Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3rd edition, 2011.
- [78] Michael K Ng, Mark Junjie Li, Joshua Zhexue Huang, and Zengyou He. On the impact of dissimilarity measure in k-modes clustering algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(3), 2007.
- [79] Robert PW Duin et al. *The dissimilarity representation for pattern recognition: foundations and applications*, volume 64. World scientific, 2005.

-
- [80] Brian S. Everitt, Sabine Landau, and Morven Leese. *Cluster Analysis*. Wiley Publishing, 4th edition, 2009.
- [81] A. Ardeshtir Goshtasby. *Similarity and Dissimilarity Measures*, pages 7–66. Springer London, London, 2012.
- [82] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA., 1967.
- [83] Glenn W Milligan and Martha C Cooper. An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2):159–179, 1985.
- [84] Jun Chen, Xuesheng Zhao, and Zhilin Li. An algorithm for the generation of voronoi diagrams on the sphere based on qtm. *Photogrammetric Engineering & Remote Sensing*, 69(1):79–89, 2003.
- [85] J Chen. A raster-based method for computing voronoi diagrams of spatial objects using dynamic distance transformation. *International Journal of Geographical Information Science*, 13(3):209–225, 1999.
- [86] Steven Fortune. A sweepline algorithm for voronoi diagrams. In *Proceedings of the second annual symposium on Computational geometry*, pages 313–322. ACM, 1986.
- [87] Der-Tsai Lee and Robert L Drysdale, III. Generalization of voronoi diagrams in the plane. *SIAM Journal on Computing*, 10(1):73–87, 1981.
- [88] Azriel Rosenfeld and John L Pfaltz. Sequential operations in digital picture processing. *Journal of the ACM (JACM)*, 13(4):471–494, 1966.
- [89] Manuel Abellanas and Belén Palop. Urban data visualization with voronoi diagrams. In *International Conference on Computational Science and Its Applications*, pages 126–136. Springer, 2008.
- [90] Michael Balzer, Oliver Deussen, and Claus Lewerentz. Voronoi treemaps for the visualization of software metrics. In *Proceedings of the 2005 ACM symposium on Software visualization*, pages 165–172. ACM, 2005.
- [91] Franck Dupuis, Jean-François Sadoc, Rémi Jullien, Borislav Angelov, and Jean-Paul Mornon. Voro3d: 3d voronoi tessellations applied to protein structures. *Bioinformatics*, 21(8):1715–1716, 2004.
- [92] Michael S Horn, Matthew Tobiasz, and Chia Shen. Visualizing biodiversity with voronoi treemaps. In *Voronoi Diagrams, 2009. ISVD’09. Sixth International Symposium on*, pages 265–270. IEEE, 2009.
- [93] Colin Ware. *Information visualization: perception for design*. Elsevier, 2012.
- [94] Ben Shneiderman. *Designing the user interface: strategies for effective human-computer interaction*. Pearson Education India, 2010.

References

- [95] Franz Aurenhammer and Otfried Schwarzkopf. A simple on-line randomized incremental algorithm for computing higher order voronoi diagrams. *International Journal of Computational Geometry & Applications*, 2(04):363–381, 1992.
- [96] Frank Dehne. An $O(n^4)$ algorithm to construct all voronoi diagrams for k nearest neighbor searching. *Automata, Languages and Programming*, pages 160–172, 1983.
- [97] Luke Bermingham and Ickjai Lee. A framework of spatio-temporal trajectory simplification methods. *International Journal of Geographical Information Science*, 31(6):1128–1153, 2017.
- [98] Atsuyuki Okabe, Barry Boots, and Kokichi Sugihara. Nearest neighbourhood operations with generalized voronoi diagrams: a review. *International Journal of Geographical Information Systems*, 8(1):43–71, 1994.
- [99] Alfred Inselberg and Bernard Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *Proceedings of the 1st conference on Visualization'90*, pages 361–378. IEEE Computer Society Press, 1990.
- [100] Almir Olivette Artero, Maria Cristina Ferreira de Oliveira, and Haim Levkowitz. Uncovering clusters in crowded parallel coordinates visualizations. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium On*, pages 81–88. IEEE, 2004.
- [101] Geoffrey Ellis and Alan Dix. Enabling automatic clutter reduction in parallel coordinate plots. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):717–724, 2006.
- [102] Helwig Hauser, Florian Ledermann, and Helmut Doleisch. Angular brushing of extended parallel coordinates. In *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on*, pages 127–130. IEEE, 2002.
- [103] Jimmy Johansson, Patric Ljung, Mikael Jern, and Matthew Cooper. Revealing structure within clustered parallel coordinates displays. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 125–132. IEEE, 2005.
- [104] Hong Zhou, Xiaoru Yuan, Huamin Qu, Weiwei Cui, and Baoquan Chen. Visual clustering in parallel coordinates. In *Computer Graphics Forum*, volume 27, pages 1047–1054. Wiley Online Library, 2008.
- [105] Gerald J Balm. Benchmarking and gap analysis: what is the next milestone? *Benchmarking for Quality Management & Technology*, 3(4):28–33, 1996.
- [106] Michael Amberg, Markus Hirschmeier, and Jens Wehrmann. The compass acceptance model for the analysis and evaluation of mobile services. *International Journal of Mobile Communications*, 2(3):248–259, 2004.
- [107] Masaaki Terashima, Ryoko Watanabe, Mikako Ueki, and Saki Matsumura. Comprehensive evaluation of antioxidant activity for various substances with 5-axe cobweb chart. *Food chemistry*, 120(1):150–155, 2010.
- [108] M Joan Saary. Radar plots: a useful way for presenting multivariate health care data. *Journal of clinical epidemiology*, 61(4):311–317, 2008.

-
- [109] Ricky KP Mok, Edmond WW Chan, and Rocky KC Chang. Measuring the quality of experience of http video streaming. In *Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on*, pages 485–492. IEEE, 2011.
 - [110] Kuan-Ta Chen, Chen-Chi Wu, Yu-Chun Chang, and Chin-Laung Lei. Quantifying qos requirements of network services: a cheat-proof framework. In *Proceedings of the second annual ACM conference on Multimedia systems*, pages 81–92. ACM, 2011.
 - [111] Yu-Chun Chang, Chi-Jui Chang, Kuan-Ta Chen, and Chin-Laung Lei. Radar chart: Scanning for satisfactory qoe in qos dimensions. *IEEE Network*, 26(4), 2012.
 - [112] Robert Tibshirani, Guenther Walther, and Trevor Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423, 2001.
 - [113] Mark Bruls, Kees Huizing, and Jarke J Van Wijk. Squarified treemaps. In *Data Visualization 2000*, pages 33–42. Springer, 2000.
 - [114] Benjamin B Bederson, Ben Shneiderman, and Martin Wattenberg. Ordered and quantum treemaps: Making effective use of 2d space to display hierarchies. *AcM Transactions on Graphics (TOG)*, 21(4):833–854, 2002.
 - [115] Ralf Dahm. Friedrich miescher and the discovery of dna. *Developmental biology*, 278(2):274–288, 2005.
 - [116] George Haddow, Jane Bullock, and Damon P Coppola. *Introduction to emergency management*. Butterworth-Heinemann, 2017.
 - [117] Guo-Dao Sun, Ying-Cai Wu, Rong-Hua Liang, and Shi-Xia Liu. A survey of visual analytics techniques and applications: State-of-the-art research and future challenges. *Journal of Computer Science and Technology*, pages 852–867, 2013.

Appendix A

Screenshots of proposed framework

This section shows screenshots of proposed research framework for few approaches introduced in this research. It illustrates outcomes that different types of HOI can be used with different VA methodologies. A user can choose different method returning various dynamic HOI. A series of these figures illustrates the interactivity aspect of proposed VA methods. With this software suite, users are able to further explore context-dependent aspects in more detail using the progressive deepening approach, and also users are able to explore different k values, different HOVDs and also diverse VA methods for “what-if” analysis. Figure A.1 displays the HOPC method with users’ interested area, it returns that HOI trends through complete k . Figure A.2 shows the HOTA method and returns the trajectory area of POI. Figure A.3 illustrates the HOS method and gives the directional HOI. Figure A.4 presents the HOR method and returns the top- k clustering HOI.

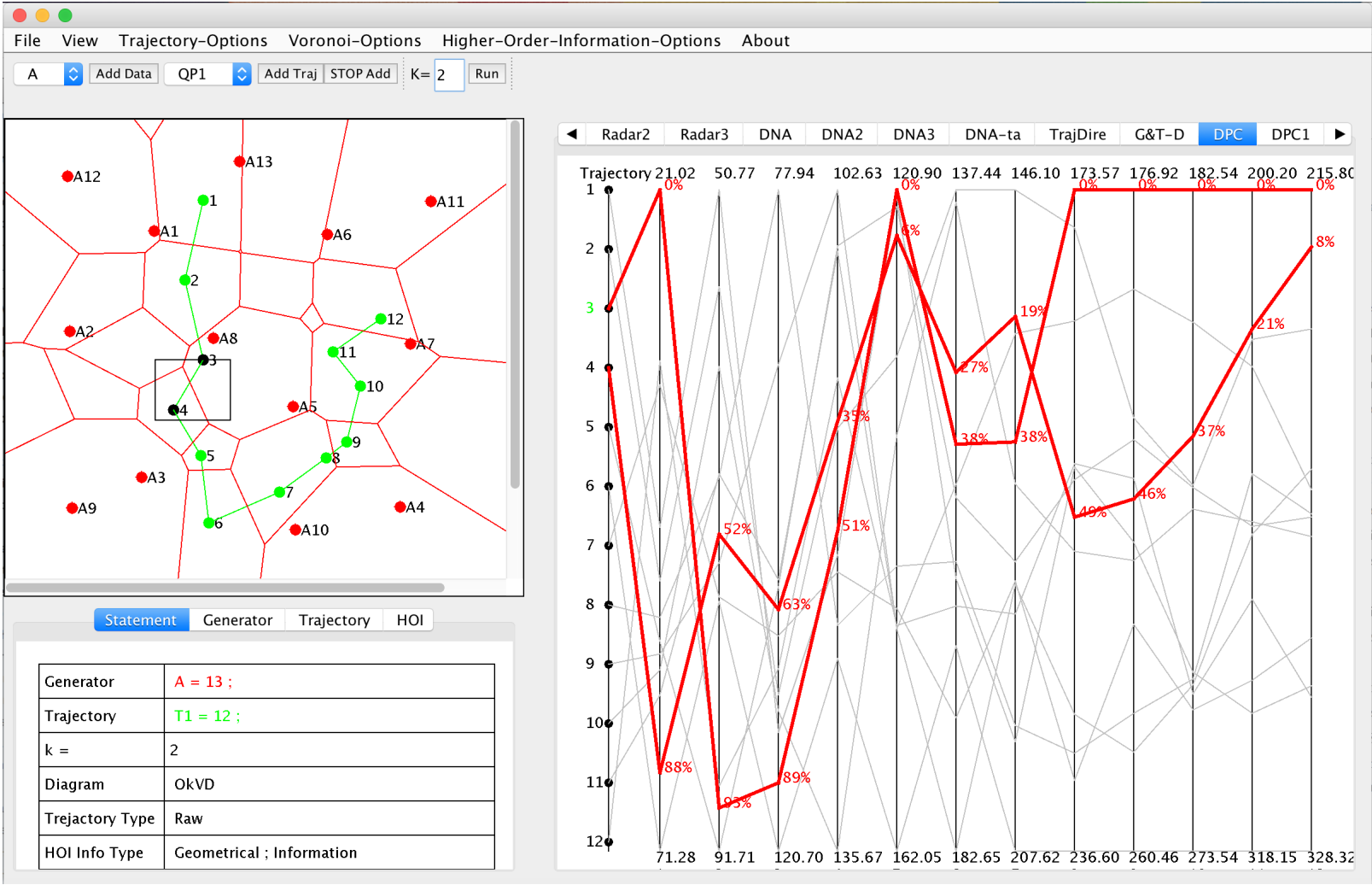


Fig. A.1 Screenshots of proposed framework for HOPC.

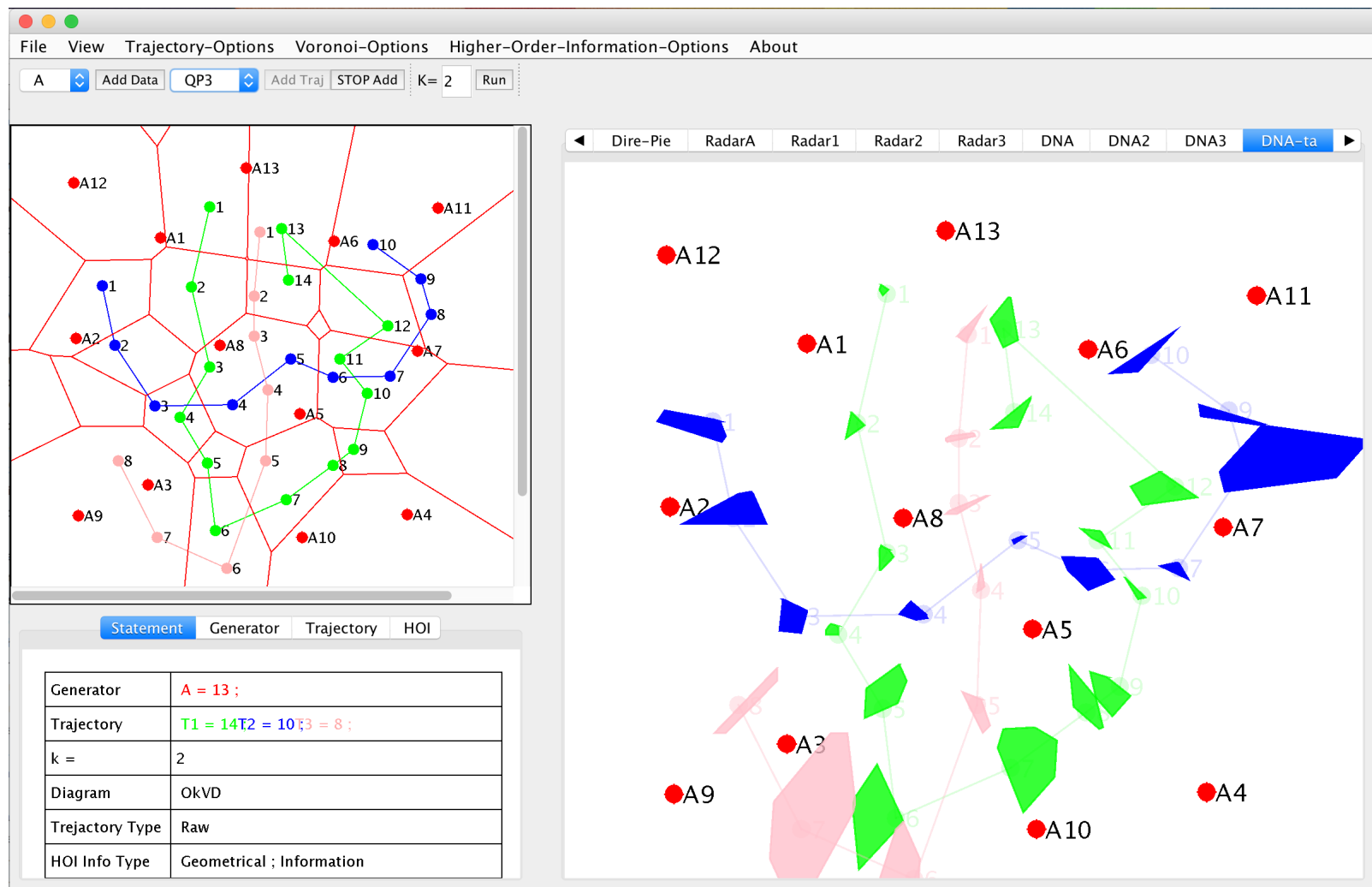


Fig. A.2 Screenshots of proposed framework for HOTA.

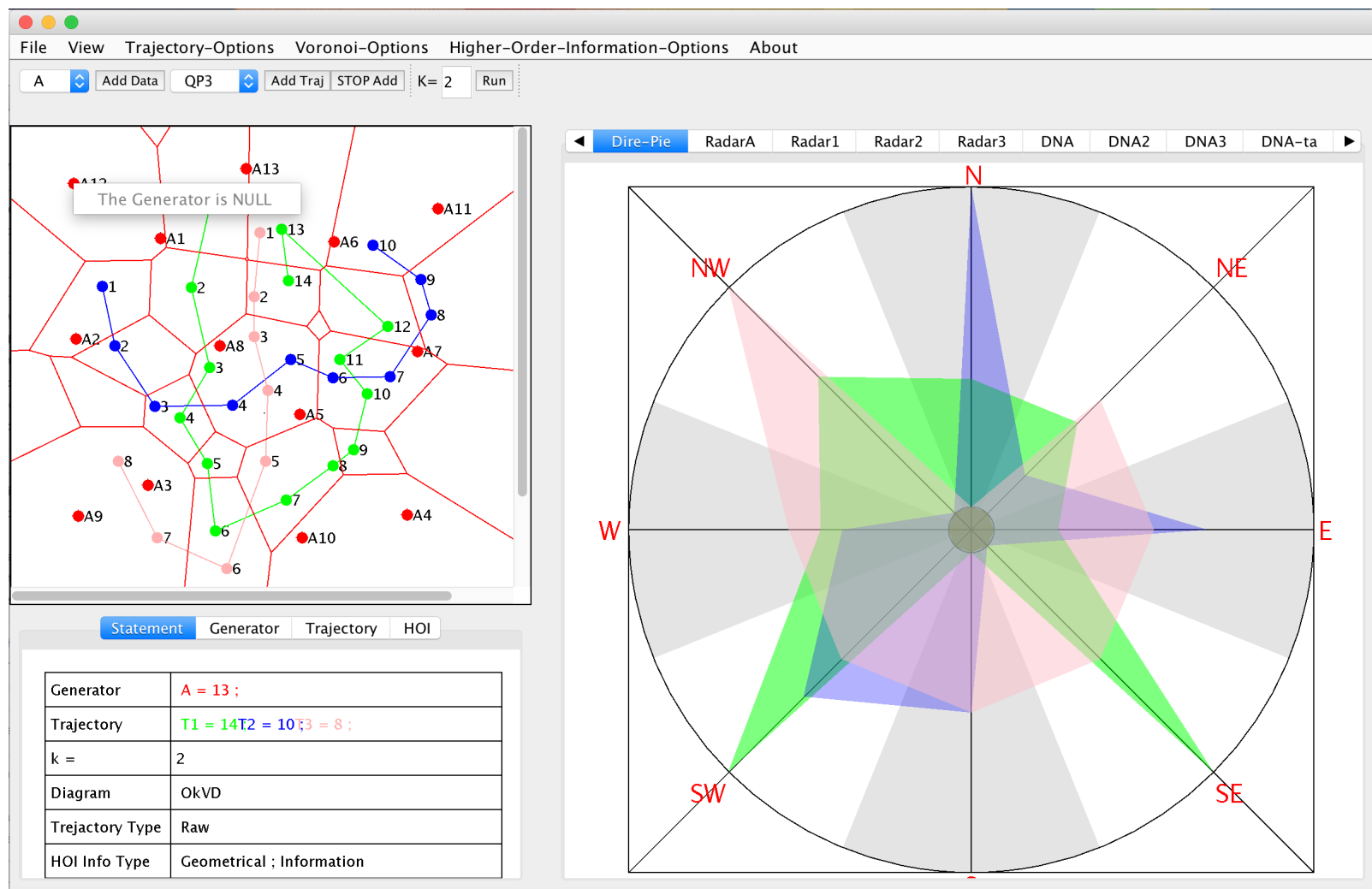


Fig. A.3 Screenshoots of proposed framework HOS.

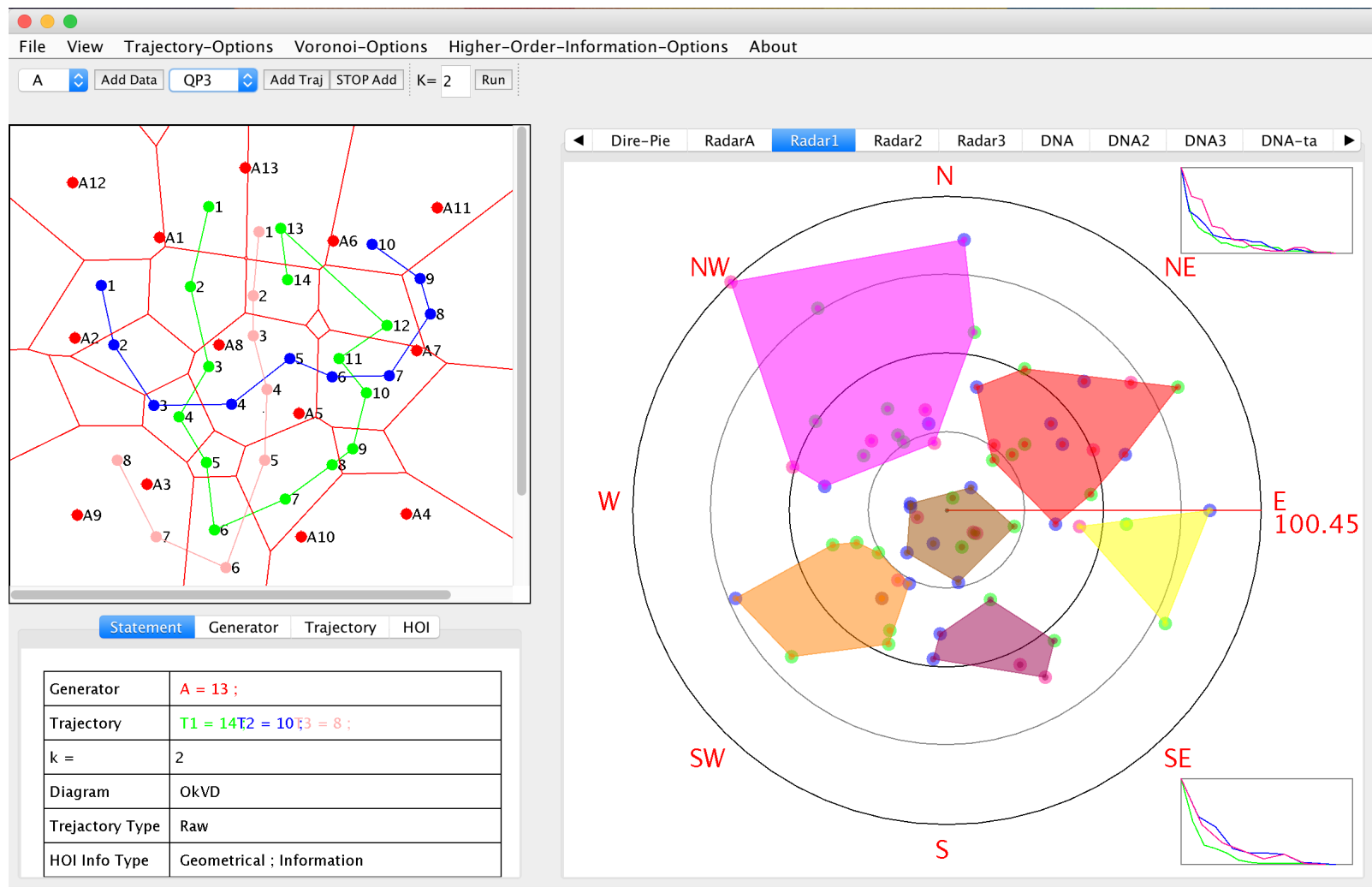


Fig. A.4 Screenshoots of proposed framework for HOR.

