Data-driven Approaches for Enhancing Resilience in Large-scale Transport Networks

Angelo Furno

Year 2023 N° to be defined

Université Claude Bernard Lyon 1

A thesis submitted in fulfillement of the requirements for the degree of

Habilitation à Diriger Des Recherches

Speciality: Informatique

by

Angelo FURNO

Data-driven Approaches for Enhancing Resilience in Large-scale Transport Networks

A journey through data analysis, traffic modelling, complex networks and flexible software architectures for resilient and sustainable cities.

Presented and publicly defended on 13 November 2023
in front of the committee composed of:

Edward Chung	Reviewer	Professor, The Hong Kong Polytechnic University, Hong Kong
Hervé Rivano	Reviewer	Professor, HDR, CITI lab, INRIA, INSA de Lyon, France
Francesco Viti	Reviewer	Associate Professor, MobiLab, University of Luxembourg, Luxembourg
Sybil Derrible	Examiner	Professor, CSUN lab, University of Illinois Chicago, US
Latifa Oukhellou	Examiner	Research Director, GRETTIA lab, University Gustave Eiffel, France
Damien Tromeur-Dervout	Examiner	Professor, Institut Camille Jordan, Université Lyon 1, CNRS, France

UMR_T9401 ENTPE - Université Gustave Eiffel

Acknowledgements

I like to imagine myself scrolling the pages of this manuscript in a few years from now, thinking of all the situations and emotions I was experiencing while trying to figure out how to weave the threads of my research into a coherent and meaningful whole.

Without a doubt, my wife and my daughter have pervaded every page of this document. Echoes of bizarre chats, kids' songs, and loud summer games will easily pop up in my mind. I feel obliged to thank them first and foremost. You have shouldered the burden of this HDR with me, tolerating my erratic moods and my mental absence while writing pages of this manuscript. Without your unwavering support, I would never have been able to complete any of these chapters. Thank you, my *beloved* Valeria and Laetitia!

Special thanks also go to my family, who have journeyed with me through my career adventures and have supported me throughout these long years despite the distance. Thank you, Carmen, Zia, Francesco, Ni & Gi, Alfio, Imma and Marco. You probably don't know, but our many remote calls provided warm comfort, helped me keep pace, and led to the completion of this work.

This HDR encompasses a constellation of work done with young and passionate PhD and postdoctoral fellows whom I have had the fortune of crossing paths with. I appreciate the excellent work you all have done. Each of you will be thanked and mentioned in this manuscript. So, please have patience once again and read the entire manuscript to find your name!

Special words of gratitude are reserved for Nour-Eddin, for whom the term mentor falls short. Your positive and generous guidance has significantly contributed to the writing of this HDR. It is a pleasure working with you!

I have also been fortunate to meet exceptional colleagues along the way: Eugenio, Marco, Razvan, and Ludovic. You have inspired me, opened my mind, and helped me achieve significant objectives in my professional life. You've also been great friends to grab a beer or two with!

A special thanks to Anne-Christine for reading and revising this document, and to all the members of the LICIT-ECO7 family for their cheerful presence in the lab.

Many thanks to the reviewers of this HDR and all the jury members for their dedication in reading this document. I hope I have managed to instil some genuine curiosity about the research I've conducted so far, and that the reading will not be overly burdensome.

A final thought goes to Salima and, more generally, to those who are no longer with me. I met you, Salima, when I first came to Lyon as a young PhD student for the SASO conference so many years ago. Who could have thought that you would support my HDR candidacy with such wonderful words and enthusiasm? You should have been present on the day of my defence, but life is sometimes so unfair. I am sure my thanks will reach you anyway, and you will share them with my mother as well.

Abstract

This Habilitation à Diriger des Recherches (HDR) thesis presents a comprehensive summary of my research activities aimed at enhancing the resilience of large-scale transport networks by joining a data-driven approach with complex network theory.

Ensuring the resilience of critical urban infrastructures, particularly the multi-modal urban transportation network, is crucial for the economic and social development of modern cities. It constitutes a significant research challenge that involves multiple disciplines, ranging from information and data science to economics and urban planning. The daily transportation of people and goods to and from urban areas relies heavily on a robust multi-modal transport network, capable of operating efficiently during normal and high-load situations, as well as adapting and recovering swiftly in the presence of recurrent and sudden disruptions. However, major cities encounter vital challenges in achieving these objectives due to urbanisation growth, extreme weather events, climate change impacts, health crises, and high pollution resulting from greenhouse gas emissions. Therefore, novel solutions are required to enhance the capability of city stakeholders to understand the actual functioning of the urban infrastructures, identify vulnerabilities and anticipate disruptions, as well as promptly implement adaptation or damage mitigation strategies following unexpected events.

My research work towards these goals is founded on three methodological pillars: *(i)* processing large and diverse datasets to gain insights into human mobility and presence practices, *(ii)* leveraging complex networks theory and methods to support real-time network monitoring, and *(iii)* developing real-time big data platforms and algorithms for resilience-oriented decision-making. The manuscript is organised into six chapters, as follows.

Chapter 1 lays the groundwork by introducing the core concepts of the manuscript and identifying the research gaps that motivated the research activities reported in subsequent chapters.

Chapter 2 delves into the potential of passively collected mobile phone data to support the understanding and management of critical infrastructures. The exploration of this data has been addressed according to two perspectives: firstly, using analytical tools to understand cellular traffic demand through the lens of call detail records, towards the efficient management and deployment of network resources; secondly, harnessing mobile phone data, and particularly network signalling data, to enhance the understanding of travel demand by accurately estimating Origin-Destination matrices.

Chapter 3 investigates further the potential of network signalling data for transportation studies in urban settings, by employing unsupervised machine learning techniques to process such data. First, it presents a framework developed to support the estimation of time-varying mean spatial speeds of multiple urban zones of a given city. Subsequently, it delves into the research efforts focused on reconstructing human mobility trajectories with unprecedented spatio-temporal accuracy, comparable to GPS data.

Chapter 4 delves into the assessment and management of vulnerabilities within complex

transport networks. This chapter describes the metrics and approaches identified to detect the most critical components of large transport networks. It also explores the design of resilient and sustainable on-demand mobility services, such as park-and-ride systems, capable of augmenting existing transportation networks with more resilient alternatives against disruptions and unexpected events.

Chapter 5 introduces the innovative real-time big data platform, PROMENADE, proposed to support the real-time monitoring and resilient management of large-scale urban infrastructures. It also presents the platform prototype tailored to road traffic networks, encompassing solutions for large-scale real-time traffic monitoring and agent-based control to reduce traffic congestion and vulnerability.

Finally, Chapter 6 concludes the HDR thesis by summarising the primary contributions and outlining short-term and long-term research perspectives that will shape future activities in the years to come.

Contents

List of Abbreviations vii

List of Figures ix

List of Tables xiii

Introduction 1
- Early Research Topics . . . 1
- Evolution of the Research Context . . . 2
- Data-driven Urban Transport Resilience . . . 3
- Structure of the Document . . . 4

1 Basic Notions 7
- 1.1 Massive Data Sources for Mobility Studies . . . 7
  - 1.1.1 Main Sources of Trajectory Data . . . 9
  - 1.1.2 Analysis of a Massive Real-world NSD Dataset . . . 11
- 1.2 Resilience Modelling . . . 14
  - 1.2.1 Main Approaches to Assessing Resilience: a High-level Classification . . . 16
  - 1.2.2 Vulnerability Analysis of Transport Networks . . . 19

2 Mobile Traffic Profiling and Travel Demand Estimation 25
- 2.1 Mobile Traffic Demand Profiling with Call Detail Records . . . 26
  - 2.1.1 Time Profiling of Mobile Traffic Demand . . . 26
  - 2.1.2 Spatial Profiling of Mobile Traffic Demand . . . 29
  - 2.1.3 Spatio-temporal Profiling of Mobile Traffic Demand . . . 33
- 2.2 Origin-Destination Flows at Regional Scale via Network Signalling Data . . . 35
  - 2.2.1 Static OD Matrix Estimation . . . 35
  - 2.2.2 Dynamic OD Matrix Estimation . . . 39

3 Zonal Traffic Speed and Urban Trajectory Reconstruction 45
- 3.1 Estimation of Mean Spatial Speeds in Urban Areas . . . 47
- 3.2 Trajectory Inference at Scale with Network Signalling Data . . . 55
  - 3.2.1 Trajectory identification . . . 58
  - 3.2.2 Trajectory Augmentation . . . 60
  - 3.2.3 Performance Evaluation . . . 61
  - 3.2.4 Large-Scale Applications of TRANSIT . . . 62

3.3 Multi-modal Path Reconstruction via Hidden Markov Model . . . 68
3.3.1 Empirical Evaluation in the Lyon Case Study . . . 71
3.4 Towards Privacy-aware Mobility Inference . . . 74

4 Assessing and Managing Vulnerabilities in Transportation Networks 81
4.1 Vulnerability Analysis: a Comparative Study . . . 83
4.1.1 Graph-theory metrics . . . 83
4.1.2 Demand-sensitive Vulnerability Metric . . . 84
4.1.3 Stress Testing Methodology . . . 85
4.1.4 Evaluation . . . 86
4.2 Graph Theory and Big Data Processing towards Online Transport Monitoring 91
4.2.1 Weighted Degree Centrality Metrics for Transport Network Resilience 92
4.2.2 Correlation of Betweenness Centrality and Traffic Metrics . . . 97
4.3 Vulnerability-aware Multi-modal Network Design . . . 100
4.3.1 Park-and-Ride for Resilient On-Demand Urban Mobility . . . 100

5 Real-time and Large-scale Solutions for Resilient Urban Networks 109
5.1 Modelling Approach and Platform Requirements . . . 110
5.1.1 Complex network modelling of city infrastructures . . . 111
5.1.2 Functional requirements . . . 113
5.1.3 Non-functional requirements . . . 113
5.1.4 Additional features . . . 114
5.2 PROMENADE architecture . . . 114
5.2.1 Live-stream and historical-data ingestion . . . 115
5.2.2 Temporary and permanent storage . . . 116
5.2.3 Stream and batch processing . . . 116
5.2.4 Service offering . . . 116
5.3 Platform implementation and its specialisation in road network monitoring . 117
5.3.1 Platform core implementation . . . 117
5.3.2 Specialisation for Road Network Resilience . . . 121
5.4 Fast Cluster-based Computation of Betweenness Centrality in Large Graphs . 125
5.4.1 Background . . . 126
5.4.2 Brandes' algorithm . . . 127
5.4.3 Equivalence class . . . 128
5.4.4 Clustering and BC computation . . . 129
5.4.5 Empirical Evaluation with Synthetic and Realistic Networks . . . 135
5.5 Multi-agent-based Large-Scale Traffic Control . . . 140

6 Conclusion and Perspectives 145
6.1 Short-term Perspectives . . . 146
6.1.1 Short-term Multi-modal Traffic Forecasting: looking for the atypical . 147
6.1.2 Synthetic Population Generation via Multi-Source Data . . . 153
6.2 Longer-term Perspectives . . . 156

Bibliography 159

List of Abbreviations

P2P Peer-to-Peer . . . 2
CDR Call Details Records . . . 3
CI Critical Infrastructure . . . 3
NSD Network Signalling Data . . . 5
OD Origin-Destination . . . 5
UAPD User Activity-Dependent Positioning Data . . . 5
GPS Global Positioning System . . . 5
IoT Internet of Things . . . 6
TAZ Transportation Analysis Zone . . . 8
FCD Floating Car Data . . . 9
RADAR RAdio Detection and Ranging . . . 8
LIDAR LIght Detection And Ranging . . . 8
BSS Bike-Sharing System . . . 9
LBSN Location-Based Social Networks . . . 9
LA Location Area . . . 11
TA Tracking Area . . . 11
GDPR General Data Protection Regulation . . . 12
CDF Cumulative Distribution Function . . . 12
PTN Public Transport Network . . . 21
TRANSIT TRAjectory inference at scale with mobile Network SIgnaling daTa . . . 21
MWS Median Week Signature . . . 30
PoIs Points of Interest . . . 30
OSM OpenStreetMap . . . 30
EFA Exploratory Factor Analysis . . . 34
EDR Enquête Déplacements Régionale (Regional Travel Survey) . . . 36
LAU Location Area Update . . . 43
MFD Macroscopic Fundamental Diagram . . . 48
MAE Mean Absolute Error . . . 55
RMSAE Root Mean Absolute Error . . . 55
MAPE Mean Absolute Percentage Error . . . 55
RMSAPE Root Mean Square Absolute Percentage Error . . . 55
HMM Hidden Markov Model . . . 69
ATG Adaptive Tree Generalisation . . . 77
BC Betweenness Centrality . . . 83

UBC Unweighted Betweenness Centrality . . . 84
TTBC Travel-Time weighted Betweenness Centrality . . . 84
STC Stress Test Criticality . . . 86
FFBC Free-Flow weighted Betweenness Centrality . . . 97
QDG Quai Dr. Gailleton . . . 99
QCB Quai Claude Bernard . . . 99
PR Park-and-Ride . . . 100
MILP Mixed-Integer Linear Programming . . . 101
LRA Lagrangian Relaxation Algorithm . . . 101
LR Lagrangian Relaxation . . . 104
SLA Service Level Agreement . . . 111
IMP Initialization and Monitoring Plane . . . 114
SI Stream data Ingestion . . . 114
HI Historical data Ingestion . . . 114
DAQ Distributed Asynchronous producer-consumer Queue . . . 114
PS Permanent Storage . . . 114
SP Stream Processing . . . 114
BP Batch Processing . . . 114
SVC Service offering layer . . . 114
SSSP Single-Source Shortest-Paths . . . 128
BFS Breadth-First Search . . . 128
AS Algorithmic Speedup . . . 135
COMFORT COoperative Multi-agent FramewOrk for large-scale Routing-based Traffic control . . . 140
MAS Multi-Agent System . . . 141
EMD Enquête Ménage Déplacements . . . 154
GNN Graph Neural Networks . . . 151
GCNN Graph Convolutional Neural Networks . . . 151
ST-ED-RMGC Spatio-Temporal Encoder-Decoder Residual Multi-Graph Convolutional network . . . 151
RMGC Residual Multi-graph Convolutional network . . . 151
MSE Mean Squared Error . . . 152
LNN Liquid Neural Networks . . . 153
EMD Enquête Ménage Déplacements . . . 154
MCMC Markov Chain Monte Carlo . . . 154
RaaS Resilience-as-a-Service . . . 156
EV Electric and electrified Vehicle . . . 158

List of Figures

1 Structure of the document: a visual representation. 4

1.1 Examples of inference of one trajectory of a volunteer from (a) CDR, and (b) NSD, compared to the corresponding GPS trace (in blue) used as ground truth. 12

1.2 CDF of inter-event times recorded in NSD, CDR, and CDR+. The plots refer to (a) median, and (b) average times per user. 13

1.3 Sample weekly trajectories of one voluntary user inferred from (a) CDR, and (b) NSD. 14

1.4 The general framework for resilience engineering proposed by Hollnagel *et al.*, 2006. 16

1.5 Transition among states of a disrupted system as proposed in Henry *et al.*, 2012. 19

1.6 Resilience curve and conditional vulnerability as from Mattsson and Jenelius, 2015. 20

2.1 Mobile traffic profile categories for the CDR of the city of Milan. 28

2.2 Residential urban fabrics. Characteristic signatures (with standard deviation) and maps of related unit areas in representative city scenarios. 31

2.3 Office fabric signatures $\mathbf{s}_2^\star$, $\mathbf{s}_5^\star$ and $\mathbf{s}_7^\star$ and maps of the related unit areas in Italian and French cities, with OpenStreetMap data. 32

2.4 Transportation fabric signatures $\mathbf{s}_9^\star$, $\mathbf{s}_{10}^\star$ and $\mathbf{s}_{36}^\star$, and their maps for French and Italian cities, with OpenStreetMap data. 33

2.5 Land use detection. (a)-(d) EFA of the total communication activity (sum of incoming/outgoing calls and SMS) in the Orange-2014 dataset. Loadings of the 1596 Voronoi cells (*i.e.,* EFA variables) on four (out of fourteen) representative classes (*i.e.,* EFA factors). (e)-(h) Thurstone's scores of the 91×24 hours (*i.e.,* EFA samples) on a selection of the 16 classes (*i.e.,* EFA factors). Figure best viewed in colours. 34

2.6 Cell tower locations (a), and (b) EDR sectors in the region of interest (Rhône-Alpes). 36

2.7 The OD matrix reconstruction framework. 37

2.8 Graphical comparison between the NSD-reconstructed OD matrix (a) and the EDR OD matrix (b). (c) Linear regression of the NSD OD matrix with respect to the EDR one. 38

2.9 Reconstruction of time-dependent OD matrices: (a) the extended framework; (b) The approach for imputing trip starting time. 40

2.10 (a) Temporal demand profile from signalling data (SD) and survey data (EDR) and demand difference between EDR and SD (red bars indicate an hourly EDR demand higher than the SD one, and blue bars otherwise) (b) correlation between the hourly demand estimations from SD and EDR. . . . 41
2.11 Analysis and correction of the biases in temporal NSD reconstructed travel demand. . . . 42
2.12 Average temporal demand profiles for a selection of clusters (a-b). Land use distribution in all clusters (e). . . . 43

3.1 Methodological framework for regional speed reconstruction. . . . 50
3.2 Maps of the regions partitioning the city of Lyon, France. . . . 54
3.3 Speed dynamics after average bias removal. Ground truth speeds from the raw GPS traces are reported as blue curves. . . . 56
3.4 Flowchart of TRANSIT. . . . 57
3.5 Main steps of the trajectory identification via TRANSIT. . . . 59
3.6 Trajectory reconstruction for a given pair of Origin-Destination path. . . . 61
3.7 Average weekly profiles of the number of concurrent trips in (a) Paris and (b) Lyon, as inferred from TRANSIT and smart card data. Normalised versions of the same profiles are presented in (c) and (d). . . . 63
3.8 Heatmap of commuting trips in Lyon and Paris. . . . 65
3.9 Typical/atypical weekly temporal demand profile during atypical events. . . . 66
3.10 Heatmap of recurrent trips for the Paris ring road (the black square shows the catchment area) . . . 68
3.11 Illustration of Hidden Markov Model based map-matching. . . . 69
3.12 Comparison between popular paths reconstructed by our approach and ground-truth popular paths for three case studies: C_1, C_2 and C_2. . . . 73
3.13 Example of generalisation and suppression of a simple OD matrix. Note that the flows here have been clustered without any kind of constraint. Some approaches seek particular solutions, notably ones where the generalisations form a partitioning of the domain. . . . 76
3.14 Example of a spatial generalisation hierarchy. The root represents the whole study area, and the children of a node form a partitioning of the parents. An individual present in area A in the data can be generalised to be shown as present in area B, C or D depending on what is necessary in order to hide them in a group of k individuals. . . . 76

4.1 Simple virtual road network. . . . 87
4.2 Stress testing on a simple test virtual network in two scenarios: performance loss for the top-5 most critical links with different capacity-disruption levels . 88
4.3 The DIRIF road network in Paris agglomeration . . . 90
4.4 The time-varying network modelling approach. . . . 93
4.5 Average speed profile, average degree centrality profile and degree distribution during two disruptive events: snow on December 18$^{\text{th}}$, 2017 (left) and tunnel closure on June 2$^{\text{nd}}$, 2018 (right). . . . 95
4.6 Analysis of disruptions in the heterogeneity-symmetry-density space. . . . 96
4.7 Correlation of per-edge average traffic flow and different BC metrics (UBC, FFBC, TTBC). . . . 98
4.8 Per-edge temporal correlation between BC and flow (a), and zoom on a specific region (b,c). . . . 98

4.9 Network segmentation (a); and the network used for the real instance I^R (b). 104
4.10 Solutions found for the real instance: green nodes are origins (potential pick-ups) and red nodes are the destinations (potential drop-offs). 107

5.1 Collecting and leveraging big smart city data as complex dynamic networks. . 112
5.2 Detailed architecture of PROMENADE. 115
5.3 Underlying software and hardware technologies of PROMENADE-v2.0. . . . 117
5.4 Detailed architecture of PROMENADE-v2.0. Gray boxes indicate that the corresponding components are not implemented in the current prototype of the platform. 119
5.5 Architecture of the IoT data-driven emulation software. 123
5.6 GUI examples - Map with vulnerability indicator (BC). The hottest spots of the heatmap represent critical nodes with high BC values. 125
5.7 Example of clustering. 130
5.8 Classes of equivalent nodes in the blue cluster C_1 130
5.9 Example of external node found through the HSN. 132
5.10 Global SSSP explorations from pivots 134
5.11 Comparison with Brandes' algorithm. Algorithmic speedup analysis - $AS^{\mathcal{E}/\mathcal{B}}_{p=[1,5,10,15,20,25]}$ 137
5.12 Comparison with Brandes' algorithm. Efficiency analysis - $E_{p=[1,5,10,15,20,25]}$. 138
5.13 Comparison with the HSN solution for synthetic (a) and real graphs (b). . . . 139
5.14 Performance analysis of the E1C-FastBC algorithm versus Brandes on the real graph lyon-road-network. 140
5.15 Activity diagram of the interactions between agents in COMFORT. 141
5.16 The northern region network of the conurbation of Lyon (3rd, 6th district of Lyon and Villeurbanne). The network contains 1,883 nodes, 3,383 links and is divided in 17 zones. Each zone is represented by a distinct colour. 143

6.1 Methodology flowchart 148
6.2 The weekly signature of counts and observations for a week in March 2019 . . 149
6.3 Average deviance of each mode per cluster (three largest identified clusters). . 150
6.4 Average deviance of each mode per cluster. 155

List of Tables

1.1 Advantages and disadvantages of novel data sources for human mobility modelling. . . . 8
1.2 Statistics on the large-scale network signalling data. . . . 12
1.3 Methods for resilience assessment, adapted and extended from Serdar *et al.*, 2022. . . . 18

3.1 Average speed errors for each mean inter-event time (without de-biasing). . . . 55
3.2 Daily speed errors with de-biasing in the worst-case scenario (20-minutes IET) 55
3.3 Descriptive statistics of the datasets used for the experiments (#matrices: number of matrices in the whole dataset, where each matrix represents the flows over a time step; #tiles: number of initial tiles over which the matrices are set; density: average graph density of the matrices; avg. #flows: average number of flows among the matrices in the dataset; avg. vol.: average sum of the flows; %anon. flows: fraction of flows that are above $k = 10$; %anon. vol: fraction of individuals that are in flows above $k = 10$.) . . . 77
3.4 Performance on samples of the dataset senegal_crop, nyc, civ, and senegal. $\bar{G}$: mean generalisation error; E: normalized reconstruction loss; S: fraction of volumes suppressed; D: distribution distance (Eq. 3.38). The reported time is the total computing time (in seconds) to run the anonymisation of all matrices of all datasets. Note that laplace_noise adds volumes, as it mostly applies a positive noise on a sparse matrix. . . . 79
3.5 Performance on samples of the datasets senegal_big and senegal_split. . . . 79

4.1 Origin-destination matrices for the simple virtual network with different demand levels (A and B). Values are expressed as vehicles per hour. . . . 87
4.2 Simple network link rankings generated by the different metrics of criticality . 89
4.3 DIRIF network link values generated by the different metrics of criticality . . 90
4.4 Results on I^R after 1 hour, with a heuristic time constraint of 20 minutes. . . 105
4.5 Park-and-ride flow shares with different budgets, using $\theta = 0.1$, $t^{\text{access}} = 2$ minutes, and $t^{\text{egress}} = 8$ minutes. . . . 106

5.1 Mapping of PROMENADE-v2.0 components to their corresponding K8s objects. . . . 118
5.2 Road segments (links) and intersections (nodes): static data provided to PROMENADE-v2.0. . . . 122
5.3 Road segments (links) and intersections (nodes): dynamic data collected and produced by PROMENADE-v2.0. . . . 122
5.4 Road segments: dynamic raw (non-aggregated) data of vehicles. . . . 124

5.5 The notation used with the BC exact computation algorithm. 127
5.6 Normalised distances and normalised number of shortest paths for the blue cluster C_1. Note that the values of $\hat{d}$ and $\hat{\sigma}$ are computed considering external node 17, as from Fig. 5.9(b). 131
5.7 Topological information of synthetic & real graphs. The names of the graphs are given in the first column, whereas the number of nodes and edges are given in the second and third columns. d_{avg} and d_{max} are the average and max degree, respectively. cc_{avg} is the average clustering coefficient. 136

Introduction

Over the last ten years, my research has progressively shifted across three major domains: service computing, analytics for cognitive 5G networks, and human mobility modelling and management. Despite this domain shift, the core of my activities has, however, remained focused on one main goal: **enabling efficient, scalable, robust, and sustainable utilisation of complex networked systems in the face of dynamic and often unpredictable *usage loads* and *contexts***. These last two terms, *usage loads* and *contexts*, may assume specific meanings according to the reference domain. From a general point of view, usage load refers to the amount of demand or activity being placed on a system by its users at any given time. Understanding and managing usage load is critical for ensuring that these systems remain stable and performant, even during periods of high or abnormal demand and in the presence of disruptive exogenous or endogenous factors. Context can be defined as the set of external factors and circumstances that affect the behaviour and performance of these systems. Analysing, modelling, and monitoring contextual factors is paramount to improve the scalability, reliability and recovery ability of the considered system, as well as to provide a better user experience.

Building on these foundations, my research journey has addressed so far a rather wide range of interrelated challenges, such as proposing solutions for data-driven understanding and modelling of user demand, performing knowledge inference from large-scale and multi-source datasets, implementing metrics, approaches and platforms for real-time monitoring of networked systems, and devising decision-making tools to anticipate or respond quickly to changes in user demand or external events. The diversity of the research topics has required acquiring expertise in several disciplines, including distributed and big data processing, complex network theory, data science and machine learning, telecommunications and transportation.

In this HDR thesis, I aim to summarise and present my main research contributions, with a particular focus on the activities from the last seven years. During this journey, I have had the opportunity to collaborate with enthusiastic colleagues who have helped me to achieve interesting findings in the aforementioned fields. I hope that this thesis adequately summarises these outcomes and fruitful collaborations.

Early Research Topics

My research experience starts during the PhD program conducted at the Engineering Department of University of Sannio, Italy, where I decided to address the problem of simplifying the lives of Web users by devising novel approaches for *automatically and efficiently composing* existing Web applications (*i.e.,* published as Web/cloud services) to generate novel value-added applications, able to satisfy complex users' needs effectively. People using the Web for digital tasks often struggle to find services that fully meet their needs directly. Instead, they have to manually and repeatedly search and join multiple services from many providers to achieve more complex tasks. For instance, when planning a long-range trip, it is necessary to identify proper applications in order to solve specific sub-tasks, *e.g.,* calcu-

lating the best itinerary to move from point A to point B; buying tickets for each involved transportation mode (*e.g.,* flight, metro, bus); finding and booking accommodation, *etc.* The general problem - referred to as *discovery by composition* - is known to be tedious when performed manually and requires high computational and time resources (*i.e.,* services are to be searched on the entire Web). Thus, my PhD thesis proposed techniques to make discovery by composition automated and scalable. In general, services and users' requirements are sensitive to their environment. The *context* is thus a crucial dimension, significantly impacting service design, query specification, and service execution. It contains implicit information essential for fully understanding users' needs. The latter are rarely properly described in computer-understandable language.

The main contribution of my PhD thesis can be summarised as follows: an integrated approach and a tool to automatically generate concrete and executable service compositions from existing Web services [1]; semantic support for enhancing context-awareness in service design, discovery and execution [2]; a fully distributed Peer-to-Peer (P2P) and cooperative approach to automatically and efficiently retrieve and join multiple services, based on epidemic diffusion protocols and semantic overlay networks [3]. We point the interested reader to [4] for more details.

Evolution of the Research Context

During the last year of my PhD thesis, my advisor at the University of Sannio, Eugenio Zimeo, invited me and other members of his research group to investigate the problem of simplifying access to digital services and resources via data-driven content recommendation. To identify targets for the recommendation task, the idea was to leverage massive data on user preferences, either explicitly declared on social media (*e.g.,* Facebook liked pages, LinkedIn groups) or implicitly determined by previous service usages performed by the user themselves or by their social environment (*e.g.,* friends, work connections, *etc.*). In pursuing such research activities in parallel with my main PhD topics, I was led to develop a new theoretical and technical background in various disciplines encompassing distributed systems and Web service technologies. Data science and unsupervised machine learning (*e.g.,* clustering) were necessary to analyse the data mined via social network APIs and identify communities of interest; information theory (*e.g.,* collaborative filtering, semantic similarity, preference mining) was required to capture the semantics of user preferences and perform item recommendation. Complex networks theory was a natural choice to model the topology of users' preferences and relationships as multi-layer networks, while social network analysis tools (*e.g.,* centrality metrics) were leveraged to spot opinion leaders within the identified interest user communities. Finally, big data processing techniques and technologies (*e.g.,* map-reduce batch programming, stream processing, Spark, NoSQL databases, such as Mongodb[1] and Neo4j[2]) became quickly indispensable to handle the huge variety, velocity and volumes, as well as the unstructured nature of the data retrieved from social networks (*e.g.,* content of liked pages, user groups descriptions, *etc.*). Based on such techniques, I contributed to laying the foundation of a social network platform for service recommendation, which was subsequently further developed by the research team at University of Sannio [5].

This experience provided me with novel knowledge and perspectives that proved crucial for my subsequent career advancement and has guided my research journey until the present day. At the end of my doctoral studies, I sought to expand my academic career internationally, delving deeper in the research related to the processing and mining of large-scale datasets, applied machine learning, big data technologies and network analysis, with a focus on incorporating greater intelligence into the urban environment for increased resilience and quality of service. This led me to apply and be selected for a CORDIS postdoctoral research fellowship at the Urbanet team, now Agora of INRIA and the CITI laboratory of INSA

[1] https://www.mongodb.com
[2] https://neo4j.com

Lyon. The postdoc was also supported by the research branch of Orange (Orange Labs, now Orange Innovation), the major telecommunication network operator in France. This marked the beginning of a strong and still-ongoing collaboration with the mobile provider, which has supported a relevant part of my research since then.

The primary contribution of the INRIA postdoc involved mining massive records of mobile phone activity, *i.e.,* mobile traffic demand, in the form of Call Details Records (CDR), to derive communication usage patterns for two main purposes: *(i)* performing data-driven resource allocation in the provider's network, aligned with the vision of 5G cognitive radio networks [6, 7, 8], and *(ii)* inferring spatio-temporal insights on the socioeconomic characteristics of the urban environment [9, 10], according to the idea that mobile phone data could provide valuable support for the comprehension of urban metabolism. This notion includes, particularly, the capacity to provide dynamic descriptions of the presence, mobility and activities of the urban population and its interactions with the urban infrastructures to support *resilient urban planning and management* [11, 12]. The most relevant research results related to these aspects are highlighted in Chapter 2.

Data-driven Urban Transport Resilience

At the end of my postdoctoral experience, I decided to join the Laboratoire d'Ingénierie Circulation Transport (LICIT-ECO7) of the University Gustave Eiffel and École Nationale des Travaux Publics de l'État (ENTPE), University of Lyon, where I am currently a permanent researcher and associate professor since October 2016. This was the opportunity of applying and extending my previous research on mobile traffic demand analysis and delve deeper into the study of data-driven resilient human mobility.

Ensuring the resilience of urban Critical Infrastructures (CIs), and particularly of the multi-modal urban transport system, is vital for the economic and social development of modern cities and represents a major research challenge involving multiple disciplines, from information and data science to economics and urban planning. The daily transport of people and goods to and from urban areas is highly dependent on a robust multi-modal transport network, capable of efficiently operating during normal and high-load situations, as well as of adapting and recovering itself quickly in the presence of recurrent and sudden perturbations. However, major cities face crucial challenges in achieving these objectives. In 2009, for the first time, more than half of the world's population lived in urban areas and by 2030, cities are expected to host 60% of the world's population. The growing concentration of people results in a continuous increase in demand for mobility, as well as frequent disruptions and significant stresses on the urban transport infrastructures. These aspects make user travel demand more uncertain and rapidly varying, degrade the mobility experience of users, deteriorate air quality and increase the risk of malfunctioning and cascading failures, especially in the presence of adverse weather conditions or extreme adverse situations. The recent COVID pandemic has exacerbated the vulnerability of the mobility system, exposing people to a high risk of contamination, deteriorated quality of service, and reduced and highly unequal access to transport.

On the other hand, cities are becoming increasingly *smart* and, therefore, able to collect vast amounts of data in real-time, via connected sensors, people and infrastructure. The recent development of 5G telecommunication networks provides reliable and ultra-fast connectivity, always-on broadband devices, and support for new computing paradigms such as fog and edge computing. In this context, massive, real-time data become available at unprecedented scales and resolution, with the great potential to provide very valuable information about how, where and when citizens move, interact, work and live, making it possible to continuously monitor and optimise the operations and usages of city infrastructures.

Based on such needs and challenges, my research activities of the last seven years at the LICIT-ECO7 have focused on the understanding of mobility practices and network

vulnerabilities, towards enhanced data-driven resilience of the transport system.

To achieve the stated objectives, my research has deliberately drawn upon previous experience, as well as methods and knowledge from multiple disciplines, including mobile phone data processing, edge/fog/cloud computing, complex network theory, machine learning methods, and big data technologies, that has flourished in a cohesive interdisciplinary approach to urban resilience engineering.

In that respect, my research revolved around the three major axes:

1. Design and development of innovative data-driven tools for the fine-grained understanding of human mobility;
2. Definition of innovative approaches, metrics and strategies for vulnerability assessment and monitoring of large-scale urban multi-modal transport networks;
3. Design and development of real-time data-driven decision tools and extensible platforms to support the resilient management of human mobility, and, more in general, of urban CIs.

As will be further detailed in the next chapters, these activities have been the result of significant collaboration with PhD and Postdoctoral students, as well as researchers from other (national or international) academic and industry institutions, that provided their invaluable help to refine and implement my research goals, methods and solutions.

Structure of the Document

This document is organised in six chapters, as visually detailed in Fig. 1.

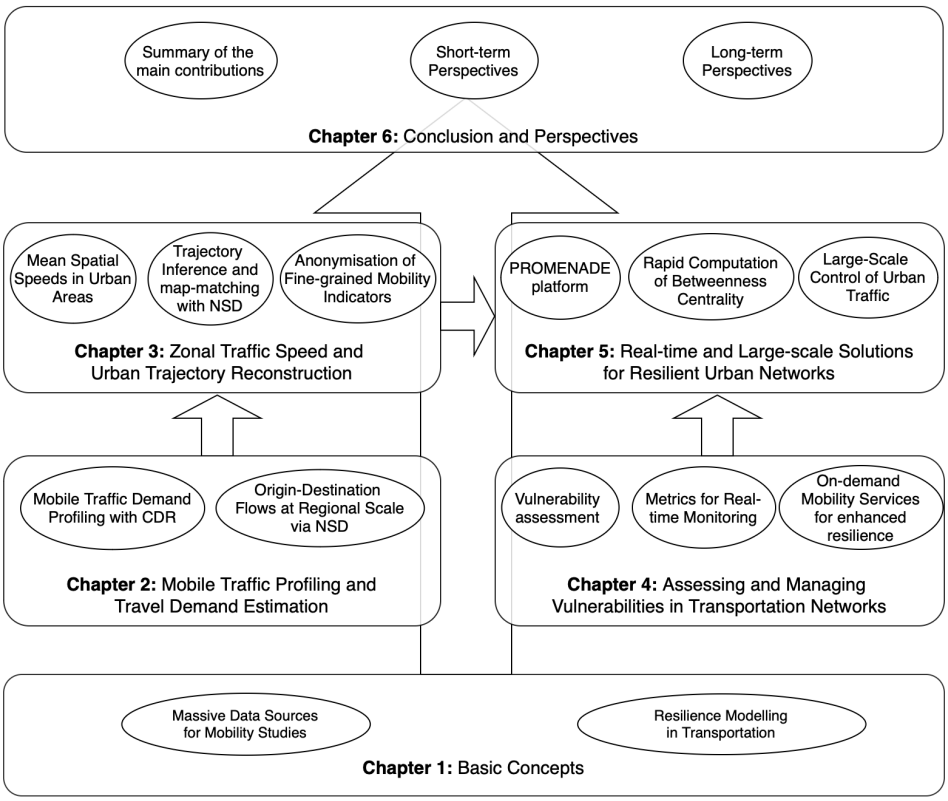


Figure 1: Structure of the document: a visual representation.

Chapter 1 prepares the ground for the rest of the manuscript by introducing the main massive data sources adopted in the reported research and by introducing the basic notions related to resilience modelling, with a special focus on the conceptual frameworks proposed by Hollnagel *et al.* [13] (*i.e., the four cornerstones of resilience*) and Bellini *et al.* [14]. The research activities described in this HDR thesis can be seen as an effort to explore the four cornerstones of resilience in an original way, *i.e.,* by utilising novel sources of massive data and combining methods from machine learning, network analysis, and big data solutions.

With this respect, the activities outlined in Chapters 2 and 3 relate to the *learn* cornerstone, *i.e., knowing what has happened*, which involves understanding via descriptive analytics how CIs of the urban system, such as the mobile phone telecommunication network and the multi-modal transportation system, are utilised by their users, in both recurring and abnormal situations.

Specifically, the first part of Chapter 2 describes the activities on the topic of CDR analytics for cognitive networking and urban fabrics detection, carried out during my postdoctoral studies. It also reports on the preliminary activities performed on Network Signalling Data (NSD) for travel demand estimation in the form of Origin-Destination (OD) matrices at a regional scale. This topic has been the subject of the PhD thesis of Mariem Fekih that I co-supervised.

Chapter 3 pushes forward the results of Chapter 2 by focusing on the reconstruction of human mobility information at a finer spatial resolution, in urban settings. First, the research activities related to mean speed estimation for urban regions are presented. To that purpose, we proposed a generic data representation, called User Activity-Dependent Positioning Data (UAPD), to model different kinds of mobile phone data. This topic was the focus of Manon Seppecher's PhD thesis, which I had the privilege of co-supervising. The second part of the chapter focuses instead on the topic of leveraging NSD to reconstruct Global Positioning System (GPS)-like human mobility trajectories in urban environment. This research was the subject of Loïc Bonnetain's PhD thesis, which I co-directed. This chapter also discusses the approaches developed to satisfy anonymisation constraints with mobile phone traces in order to ensure confidentiality in the collection, processing and sharing of mobility indicators derived from NSD. This subject is one of the topics of another ongoing PhD thesis (Benoit Matet) co-advised with the GRETTIA laboratory of University Gustave Eiffel.

Chapter 4 outlines the research efforts focused on modelling, analysing, and managing the vulnerabilities of multi-modal transport networks. First, we present research activities centred on the cornerstone of resilience known as *anticipating*, *i.e., knowing what to expect*. These activities have been helpful in highlighting the limitations of current approaches for identifying vulnerabilities to disruptions in complex and highly dynamic transportation networks. Building on these insights, the chapter explores the combined use of complex network metrics and mobility data for continuous monitoring. This approach enables quick detection and anticipation of disruptive events in large-scale settings. Our methodology offers a dynamic and reactive approach to the resilient management of transportation networks. It provides advantages over traditional vulnerability assessment methods by better accounting for events that are hard to model and predict. This is in line with the *monitoring*, *i.e., knowing what to look for*, aspect of resilience. Finally, we present an urban planning strategy aimed at integrating multi-modality, specifically focusing on on-demand shuttle services. This strategy aims to reduce the vulnerability of specific urban areas to recurrent disruptions while preserving efficient network operations under typical conditions. This aligns with the *responding*, *i.e., knowing what to do*, aspect of resilience. The research reported in this chapter includes significant contributions from Elise Henry's PhD thesis, which I co-supervised.

Chapter 5 expands upon the findings of Chapters 3 and 4. It introduces the architectural, technological, and algorithmic support developed to support the effective processing of both large historical and real-time stream data. Moreover, it discusses the strategies

implemented to address the two core elements of resilience introduced in the previous chapter: continuous *monitoring*, *i.e., knowing what to look for*, and *responding*, *i.e., knowing what to do*. Specifically, the first part of the chapter focuses on the design and prototyping of the PROMENADE general-purpose platform, its prototype for road networks and related applications. The platform is part of the homonymous project, ANR JCJC PROMENADE, which I have led from February 2019. Inspired by the principles of the Internet of Things (IoT), cloud computing, and micro-services architectures, the platform supports the continuous development of data-intensive and real-time applications. The second part of the chapter delves with the applications of the platform prototype. A first one leverages a betweenness-centrality-based approach for quasi-real-time monitoring of very large networks. This application exploits a new algorithm for the rapid calculation of the betweenness centrality indicator on large networks, which is also described in this chapter. A second application focuses on the large-scale control of urban road traffic through a multi-agent cooperative framework. Both applications were developed in the context of Cecile Daniel's PhD thesis, which I co-directed in collaboration with the LIRIS laboratory of University Claude Bernard Lyon 1, and in collaboration with the University of Sannio.

The final Chapter 6 summarises the primary contributions of my research activities thus far and highlights future research perspectives that I will be pursuing in the short and long terms.

Chapter 1

Basic Notions

This chapter introduces the foundational notions underpinning this manuscript, organised into two primary sections.

In the first section, a succinct overview of the most relevant large-scale data sources used in mobility studies is presented. The focus is solely on the data incorporated into my research, with no intention of exhaustively covering the vast field of data-driven transportation studies. This section also includes a comparative analysis of NSD and CDR data, based on a real-world dataset provided by Orange Innovation. The concepts introduced in this section represent the necessary groundwork for the rest of the manuscript, particularly for Chapters 2 and 3, which focus on the analysis and processing of mobile phone data.

The second section shifts to resilience modelling, introducing specific vocabulary to assist the reader in accurately interpreting the manuscript when discussing resilience and related concepts. Considering the broad nature of resilience, rooted in the fields of critical infrastructures and risk management, this section delves deeper into its nuances as it pertains to transportation. Particularly, this section provides a thorough exploration of related work on vulnerability assessment for multi-modal transport networks, and identifies the most pertinent research gaps that motivated the activities described in Chapters 4 and 5.

1.1 Massive Data Sources for Mobility Studies

A key objective of my research at LICIT-ECO7 has been to investigate the potential of large-scale mobility data sources in order to develop automatic approaches to derive insights on human mobility. This exploration has contributed to enriching traditional modelling of travel demand and enhancing the management of transport networks. In that respect, a strong attention has been reserved to data sources preserving trajectory information at an individual level, and, particularly, to mobile phone passive data. Working on such data has come with various challenges, such as overcoming the inherent limitations in terms of accuracy of these data sources, which are not explicitly thought to provide mobility information, as well as identifying proper combinations of techniques and technologies to enable the automatic processing of their huge volumes. In the following, the most relevant sources of massive trajectory data are shortly (and definitely non-exhaustively) described, together with a brief analysis of their main advantages and limitations (see Table 1.1). The interested reader can refer to recent surveys on the topic, such as [15, 16, 17], for a more detailed bibliography on these data sources and their applications in the field of human mobility study and modelling.

For decades, household surveys, completed through phones, the Internet or face-to-face interviews, have been the major source of data for analysing and modelling travel demand at city, regional or even nation-wide scales. Actively compiled by travellers (or by operators

Table 1.1: Advantages and disadvantages of novel data sources for human mobility modelling.

Data Source	Advantages	Disadvantages	Related Chapters
GPS	• High spatial accuracy and temporal resolution • Speed data available	• Privacy concerns • Small sample size • Unavailability of data (*e.g.,* tunnels)	• Chapter 3 • Chapter 4 • Chapter 5 • Chapter 6
Ticketing and smart-card	• Large sample size • Low collection cost	• Limited to transit users • Incomplete trip records (*e.g.,* missing tap-out)	• Chapter 3 • Chapter 6
BSS Data	• Detailed urban mobility patterns • Reflects infrastructure utilisation and behavioural shifts • Low collection cost	• Limited to users of the specific BSS • Potential privacy concerns • Data might be skewed in areas with lower BSS penetration and affected by exogenous factors	• Chapter 6
LSBN	• Can provides detailed information on trip motifs (*e.g.,* leisure, sport) and some mode choices	• Limited geographic coverage • Data accuracy issues	• Chapter 3, in the form of UAPD
Call Detail Records	• Very large sample size • Low collection cost	• Limited to telecommunication network users • Spatial resolution limited to antenna's coverage area • Limited temporal resolution	• Chapter 2
Network Signalling Data	• Very large sample size • Low collection cost • High temporal granularity	• Privacy concerns • Noise and oscillation effect	• Chapter 3 • Chapter 4

on their behalf) in the form of travel diaries, they have fed plenty of transportation studies aimed to reconstruct mobility indicators, *i.e.,* the daily number of trips, travel mode choices, routes, motifs and typical travel times, usually defined between OD pairs of Transportation Analysis Zones (TAZs).

However, as will be further discussed in Sec. 2.2, household travel surveys suffer from multiple well-known issues [18]: they can be very expensive to conduct, become quickly outdated, and, due to high non-response rates, are unavoidably biased towards the relatively small part of the respondent population.

Other relevant sources of mobility data include loop detectors, camera, RAdio Detection and Ranging (RADAR) and LIght Detection And Ranging (LIDAR) sensors, which usually only provide aggregate counts of moving objects, thus not allowing to follow their movements in space and time.

Triggered by new technologies, this situation has deeply changed with the emergence of novel (often said *opportunistic*) sources of massive data for transportation. These data can be collected via, among others, smart ticketing public transit systems, widely adopted navigation systems and mobile applications (such as TomTom, Google Maps, Waze, *etc.*), highly popular social media platforms, or mobile phone records collected by telecommunication service providers. All these sources offer advanced possibilities to *passively* collect information on large masses of travellers and, consequently, to extract large-scale spatio-temporal patterns related to multi-modal travel demand and traffic variables.

1.1.1 Main Sources of Trajectory Data

Among the most traditional sources of trajectory data, GPS logs are produced using information from multiple satellites that provide precise localisation of moving objects (like smartphones, vehicles) equipped with a GPS receiver. Modern navigation systems can thus use the GPS to record the trajectories of a users movement with a high degree of spatial accuracy (around 5 meters) and high temporal resolution (in the order of a few seconds), regularly uploading the collected data to the navigation service provider's back-end system. Floating Car Data (FCD) refers to the collection of traffic and mobility information from moving vehicles, predominantly leveraging the GPS technology. In essence, vehicles act as mobile sensors on the road, transmitting data about their speed, direction, and location. Even though GPS-derived data could be very accurate, it is rare to achieve a high penetration rate in the travellers' population, which makes it complicated, if not impossible, to perform at affordable costs city-wide or region-wide studies that require, *e.g.,* the estimation of vehicle flows. However, FCD remain an invaluable data source for accurately analysing mobility trajectories and estimating and forecasting congestion states, especially at local scales like road segments or areas crossed by heavily trafficked roads. Due to these properties, FCD are often used as a source of ground truth to measure the quality of reconstruction methods that utilise other datasets (see Chapters 2 and 3). Moreover, FCD can be instrumental in identifying anomalies and responses to disruptions by comparing traffic conditions during specific events to typical situations averaged over multiple days of observations (see Chapter 4). GPS applications for mobility are extensively surveyed in [19].

Public transport authorities usually equip their transportation infrastructure (*e.g.,* metro and bus stations) and means (*e.g.,* buses, tramways) with ticketing data collection systems, which allow storing aggregate (*e.g.,* hourly) and spatially localised counts of transit usage. Recent public transit ticketing systems allow travellers to carry a personal smart-card, or their own credit card, and check-in/check-out for boarding and alighting at public transport stops. These systems can thus generate, at large scale, massive public transport trajectory data including, *i.e.,* card identifier, stop origin, boarding time, stop destination, alighting time for each individual transit user. Smart-card data have thus the advantage to provide a large coverage of public transport trips, even at an individual level, with very high accuracy both in space and time (see Chapter 3 for their use to validate aggregate statistics on mobility derived from NSD). Nevertheless, the data suffers from missing validations (faults, frauds, *etc.*), and cannot provide any information on other transportation modes such as car, bike, walk, train. A detailed analysis of mobility patterns detection via smart card data can be found in [20].

As a sustainable alternative to motorised vehicles and transit, Bike-Sharing Systems (BSSs) serve as a crucial data source from a research perspective, shedding light on active mobility patterns, infrastructure utilisation, and evolving human behaviours. Typically, BSS datasets capture trip duration, start and end points, and often include user demographics. Analysing BSS data allows cities to appreciate the role of cycling as a resilient mode of urban mobility, aligning with current health challenges and broader sustainable development goals. Systems like Velo'v in Lyon and Vélib in Paris have evidenced the transformative potential of BSSs, as demonstrated by the surge in cycling trips following their launches [21, 22]. The COVID-19 pandemic further underscored the significance of BSSs in steering urban mobility transformations rooted in ecological awareness [23]. With public support, infrastructure, and incentives, BSSs advocate for a transition to sustainable, multi-modal mobility, especially in urban environments. However, it's worth noting that while BSSs provide a convenient commuting option, cycling is still influenced by external factors like weather, topography, and urban design. These considerations present research challenges that must be acknowledged when leveraging BSS data in transportation studies (see, *e.g.,* Chapter 6). A comprehensive review on this topic can be found in [24].

Location-Based Social Networks (LBSN) data have more recently emerged with the widespread diffusion of social networking applications such as Twitter (now X), Facebook,

Instagram, Strava. These data typically include temporal and localisation information, together with content information (text, photos, ratings, *etc.*) related to the interaction of people with places of the urban environment (*e.g.,* check-ins at restaurants or museums, sport activities in parks, participation to concerts and public events, *etc.*). LBSN data can thus enable the identification of spatio-temporal patterns related to mobility, socioeconomic urban activities as well as attendance and perception of places. Despite such advantages, the availability of location data with LBSN is only limited to a small percentage of the social networking population and app usages, which again hampers the possibility of performing statistically relevant and large-scale studies on human mobility. A detailed survey on usages of LBSN data for urban studies is provided in [25].

As already discussed in the introduction to this manuscript, CDR issued from mobile network providers deserve a special attention as an indirect source of mobility data. CDR are data collected by mobile network operators that contain information about the usage of mobile phone services by customers. This data typically reports the date and time of call or text message initiation and termination, the numbers of the calling and called party (before anonymisation), the duration of the call or message, and the location of the phone when the call or message took place [26]. The location of a Call Details Records log is usually provided in terms of the identifier of the antenna where the communication event has taken place. In recent data collections, CDR typically also report information on data connections and the amount of traffic downloaded/uploaded by mobile phone apps running on the phone. For privacy concerns, CDR are typically pseudo-anonymised[1] and/or aggregated in time and space. Multiple CDR logs related to different users are summarised by reporting the total number of observed events at the given base station, with a given time periodicity (*e.g.,* every hour). This is the kind of data used in the research activities described in Sec. 2.1.

CDR present a unique combination of desirable properties: (*i*) they offer unprecedented penetration as they are available for the whole subscriber base of a network provider, which typically covers tens or hundreds of millions of users; (*ii*) they are recorded continuously over long time periods, allowing fine-grained longitudinal studies over months or years; and, (*iii*) they are passively collected and maintained in curated databases for billing purposes, which makes them a very cost-efficient source of data for secondary use and analysis.

Due to such interesting properties, CDR have become an important source of data for various fields of research, such as telecommunications, urban planning, epidemiology, and sociology. In telecommunications, CDR have been used to study patterns of communication, the evolution of social networks, and the impact of mobile phone services on the economy. In transportation and urban planning, CDR have been employed to derive and validate general laws that govern human movements [27], reconstructing OD matrices [28], understanding urban land use dynamics [9, 10] or inferring population density shifts over time [29]. In epidemiology, CDR have been used to study the spread of infectious diseases [30]. In sociology, CDR have been used to study social inequalities, community structures, and individual behaviours [31]. In this sense, CDR are a primary source of information to *learn* the behaviours of urban complex systems and CIs. Results related to the use of aggregate CDR data are reported in the first part of Chapter 2 (Sec. 2.1).

However, and despite their significant advantages for human-centric mobility studies, CDR have fundamental limitations in terms of positioning accuracy in both space and time. In space, the mobile device locations can only be mapped to the coverage area or position of the base stations to which it is associated [32]; in time, the sampling process is driven by the occurrence of voice call establishments or text message transmission, which are both sparse and irregularly distributed [33]. Lastly, even though they might cover a significant portion of the population depending on the operator's market share, these data are inherently skewed by the socio-economic traits of the provider's network user base. Ultimately, these problems limit the utility of CDR for studies that require a high level of spatio-temporal detail [34], making it challenging to gain a detailed understanding of such behaviours.

[1] The identifiers of the caller and the callee are replaced by random identifiers.

NSD represent another kind of mobile phone passive data, which allows to overcome such limitations. NSD can be considered the latest generation of CDR-like data. Compared to CDR, NSD convey richer information on the activity of a mobile phone user, and can be collected for the entire subscriber base of a mobile telecommunication provider. They include all the network data-plane events generated by every device associated with the provider's radio access network across 2G, 3G, 4G and, more recently, 5G cellular technologies. NSD events are triggered by a variety of interactions: (*i*) voice and texting communications (*i.e.,* call establishments and SMS transmissions, which are fully equivalent to those logged by CDR), (*ii*) handovers (*i.e.,* device cell changes during communication), (*iii*) Location Area (LA) and Tracking Area (TA) updates (*i.e.,* cell changes that cross boundaries among larger regions named LA in 2G/3G and TA in 4G, also affecting idle devices), (*iv*) active paging (*i.e.,* periodic requests to update the location of the device started from the network side), (*v*) network attaches and detaches (*i.e.,* devices joining or leaving the network as they are turned on/off), and (*vi*) data connections (*i.e.,* requests to assign resources for traffic generated by mobile applications running on the device).

The diversity of events encompassed by NSD inherently enriches the data with additional information and increases their temporal frequency compared to CDR. Due to such properties, these data are appealing to explore human mobility at scales impossible to analyse with alternative data sources traditionally used in transportation, such as household surveys or GPS data. However, NSD come with fundamental issues and limitations, such as low accuracy in space, sparsity in time and high noise due to oscillation effects. These issues make them not directly applicable for fine-grained mobility studies. Analysing the potential of such data and addressing their issues for mobility analysis have been at the centre of the research described in the second part of Chapter 2 (Sec. 2.2), which presents the results of the initial research conducted on test instances of NSD. Despite the limitations associated with these initial data, such activities allowed verifying the potential of novel generation mobile phone data for reconstructing travel demand at a regional scale.

As further elaborated in Chapter 3, these attributes also make the data substantially larger in size, and more prone to noise, manifested as higher oscillation occurrences, than what is seen in traditional CDR data. Moreover, the characteristics of NSD don't necessarily enhance the spatial resolution of each logged event, which is still localised at the level of a network base station. Consequently, more sophisticated machine-learning based processing techniques are imperative to transform this raw data into meaningful and fine-grained insights regarding human mobility. Utilising extensive individual mobile phone datasets, Chapter 3 demonstrates the capability to derive detailed speed estimations (Sec. 3.1), GPS-like trajectories (Sec. 3.2) and multi-modal routes within the transportation network (Sec. 3.3). It also raises considerations about the need for anonymisation methods, and suggests strategies to retain highly-detailed mobility information while upholding privacy constraints (Sec. 3.4).

1.1.2 Analysis of a Massive Real-world NSD Dataset

To provide a basic understanding of how mobile phone data can provide useful information on mobility trajectories, this section introduces a real-world large-scale NSD dataset provided by Orange Innovation in the context of the ANR PROMENADE project, and related to the cities of Paris and Lyon. This section also provides empirical evidence of the increased accuracy of NSD when compared to CDR. This dataset has supported most of the research studies described in Chapter 3.

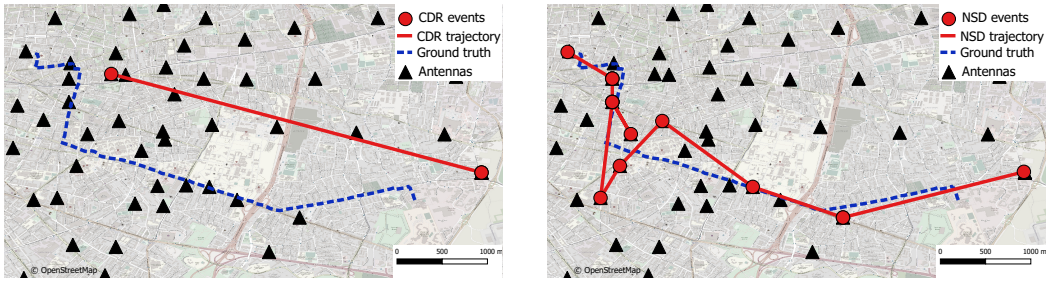
Firstly, we separated the dataset in two parts, denoted as $\mathcal{D}_P$ and $\mathcal{D}_L$, for Paris and Lyon, respectively. $\mathcal{D}_P$ and $\mathcal{D}_L$ were gathered during three consecutive months in 2019, from March 15 to June 15. The data collection, performed via the operator's network probes, was authorised by the Data Protection Officer (DPO) of Orange according to article

89 of the General Data Protection Regulation (GDPR)[2], which provides an exemption for research, in particular for scientific and research purposes. The data were collected and processed exclusively on the Orange Innovation secure Big Data platform. The data were also pseudonymised and stored in a private directory in a server located in the operator premises, and accessible only to authorised researchers. All source data were deleted 12 months after the collection. The datasets included more than 150 billions of logged events overall, observed on a mobile phone network including more than 4,600 antennas. The resulting total user base tallied to over 10 millions of individual mobile subscribers identifiers (IMSI) and over 3 millions of estimated residents in the two considered cities. Additional details on the $\mathcal{D}_P$ and $\mathcal{D}_L$ NSD datasets are reported in Tab. 1.2.

Table 1.2: Statistics on the large-scale network signalling data.

Dataset	City	Area (km^2)	Nb antennas	2G events ($\cdot 10^6$)		3G events ($\cdot 10^6$)		4G events ($\cdot 10^6$)	
				Nb IMSI	Nb events	Nb IMSI	Nb events	Nb IMSI	Nb events
$\mathcal{D}_L$	Lyon	1,506	646	1.7	83	2.8	1,470	2.9	20,994
$\mathcal{D}_P$	Paris	5,784	3,972	5.9	850	6.5	10,166	6.1	116,461

The main challenge we addressed consisted in reconstructing from $\mathcal{D}_P$ and $\mathcal{D}_L$ the separate trips of each mobile phone device with a high spatial accuracy to infer information about the exact paths travelled (*e.g.,* as the sequence of road segments, or the combination of transport modes). City environments exacerbate the problem, as they feature difficult-to-track short trips over entangled dense road layouts with multiple transportation modes. Traditional CDR are not suitable to address the task, due to their limited spatial resolution and sampling frequency. For instance, Fig. 1.1(a) shows the localisation samples recorded by CDR for an exemplary urban displacement; a linear interpolation of the CDR samples (solid red) is superposed to the same user trajectory, as observed via GPS data (dotted blue). The figure makes it clear that inferring the actual movement from CDR is an arduous mission.



(a) Trajectory from Call Detail Record

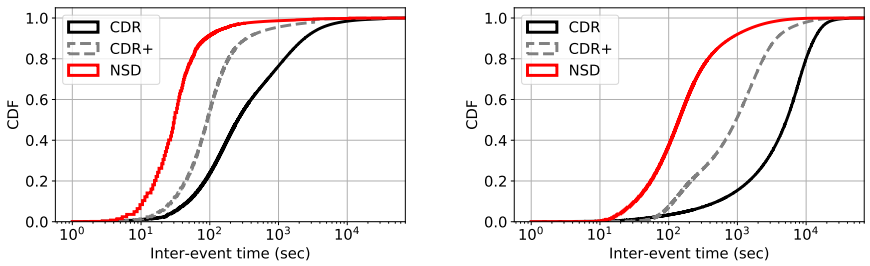
(b) Trajectory from Network Signalling Data

Figure 1.1: Examples of inference of one trajectory of a volunteer from (a) CDR, and (b) NSD, compared to the corresponding GPS trace (in blue) used as ground truth.

By visually comparing Fig. 1.1(b) to Fig. 1.1(a), it appears evident how the variety of collected events with NSD may increase significantly the number of points observed for a given user, paving the way to finer-grained studies of human mobility. The assortment of situations captured by NSD (*i.e.,* voice calls, text messages, handovers, location updates, active paging, attaches/detaches and data connections) is much wider than the sole call- and text-related events in CDR. This naturally leads to a much higher sampling frequency of the locations of devices (hence, users) over time in NSD with respect to traditional CDR.

A quantitative inspection of the increased temporal accuracy of NSD is provided in Fig. 1.2. The two plots present the Cumulative Distribution Function (CDF) of the time

[2] https://gdpr.eu/tag/gdpr/



(a) Median (b) Average

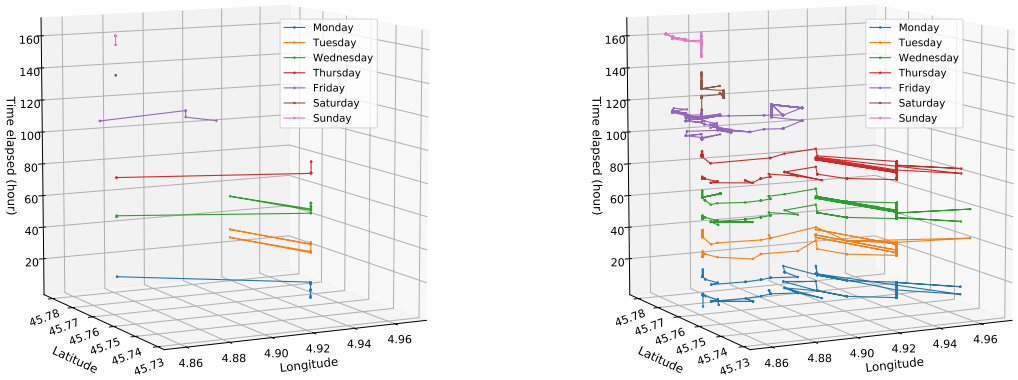
Figure 1.2: CDF of inter-event times recorded in NSD, CDR, and CDR+. The plots refer to (a) median, and (b) average times per user.

between subsequent NSD events; specifically, the distributions are computed over the (a) median and (b) mean inter-event time recorded for each device, hence they provide a fair view of the statistics across the observed population. We also report equivalent CDF obtained using other kinds of mobile network data: (*i*) CDR, which, as already mentioned, only capture voice and texting communication events, and (*ii*) CDR augmented with LA and TA update events, which we term CDR+. The rationale of this comparison is that CDR are the most widely adopted source of data from mobile networks, whereas CDR+ have been recently used for human mobility trajectory inference in some studies from the related literature [35]. We directly extrapolated CDR and CDR+ from the available NSD database, by simply retaining only the spatio-temporal samples generated by the events that are captured by such data sources, while filtering out the information associated to all other network event types.

The distributions in Fig. 1.2 yield a number of interesting observations. NSD grants a median inter-event time below 1 minute for 90% of the users, while that figure grows to 5 minutes for CDR+ and over 30 minutes for CDR. Per-user averages that are biased by long inactivity periods highlight even more the difference between the data sources: NSD keeps averages below 15 minutes for 90% of the users, whereas CDR+ and CDR record mean inter-arrivals of up to 1 hour and 3.5 hours for the same user fraction. The conclusion is that NSD ensure a sampling rate increase of more than one order of magnitude with respect to CDR and of a factor 5 over CDR+. Importantly, these results are fairly uniform over the considered population.

NSD are not supposed to bring any advantage over other classes of mobile network positioning data in terms of the absolute spatial accuracy *of each location sample*. As a matter of fact, NSD, CDR, CDR+, and any other network data types, are collected on the same radio access network infrastructure: therefore, the locations used to geo-reference the events are those of a matching set of base stations to which mobile devices associate over time. To prove our point, we run experiments with ground truth GPS data collected by a small set of volunteers. For each volunteer, we computed the distance between the location of the antenna associated to all generated network events and the corresponding GPS position at the time. Repeating the process for all CDR, CDR+ and NSD events yielded very similar average distances, between 0.26 and 0.28 km, in the three cases.

However, NSD provide a much more accurate spatial representation *of the trajectory as a whole*, as a direct consequence of the increased sampling rate. This is clearly shown in plots (a) and (b) of Fig. 1.1 for a single trajectory, as well as in plots (a) and (b) of Fig. 1.3 for multiple trips of a same user. These figures highlight the capability of NSD to capture individual mobility patterns in a much more exhaustive way compared to CDR. The unprecedented spatio-temporal resolution of NSD is at the basis of the framework described in Sec. 3.2.



(a) CDR (b) NSD

Figure 1.3: Sample weekly trajectories of one voluntary user inferred from (a) CDR, and (b) NSD.

An important aspect of the data employed for our study is that it covered three generations of cellular network technologies. This let us investigate the relevance of events generated by 2G, 3G, and 4G events on the accuracy of the positioning data. Tab. 1.2 breaks down the number of unique devices observed under each technology, as well as the number of events recorded, separately reported for the two large-scale datasets related to Paris and Lyon, $\mathcal{D}_P$ and $\mathcal{D}_L$, respectively. The figures evidence how the number of users that can be monitored by the three radio access technologies is comparable, and partially overlapping. However, the sets of geo-referenced NSD collected for the monitored devices is completely different: the number of events grows by more than one order of magnitude when moving from one cellular generation to the next.

While this is a clear result of the increased consumption of mobile services and associated growth of mobile data traffic that newer network technologies support, it further distinguishes our study from the many previous works that date back to the 2005-2015 period, and that could only rely on limited 2G and 3G data (which includes also the results discussed in Sec. 2.2).

1.2 Resilience Modelling

To guide readers through the vast and often misapplied terminology surrounding "resilience", a word already frequently used in this document without a precise definition, we present a brief introduction below. It covers essential terms, definitions, and methodologies from the realm of resilience modelling, which will be employed throughout this document

The word "resilience" is derived from the Latin term *resilire*, meaning "to bounce" or "to rebound". The term was initially used in material science to describe the resistance of materials to physical shocks and has been since then widely used in other fields such as biology [36], social sciences [37], and engineering [13] to denote the speed at which a complex system bounces back from the degradation of its functions deriving from disruptive events [38]. Resilience has also important relationships and differences with the concept of *reliability*, the latter usually referring to the ability of a system to consistently perform its functions for the period of time intended under *a well-specified set of operating conditions encountered* [39]. Moreover, reliability typically assumes no or weak correlations among the different system components that can exhibit failures as a consequence of a disruption. Conversely, resilience does not consider explicit assumptions on the nature and extent of disruptions, which often have the nature of *surprises* [40], and presumes the existence of possible complex interactions and inter-dependencies between the system components,

forming a *complex network* [38]. These differences are particularly relevant in our context of transportation networks, exposed to hardly predictable disruptive events and composed of multiple topologies associated with different transport modes, typically interacting according to complex spatio-temporal patterns, and often exhibiting long-range spatial correlations [38].

Given the diverse and complex responses of systems to perturbations, as well as the highly heterogeneous nature of the perturbations, resilience is a multifaceted and broad domain of research. As a result, there is currently no consensus on the definition of resilience, as noted in prior work [41]. First of all, as discussed in [42], resilience can be defined with respect to four dimensions: *technical*, *organisational*, *social*, and *economic*. Technical resilience relates to the ability of physical systems to perform well during disruptions through the engineered features of the constituting and interacting components. Organisational resilience refers to the capability of critical facilities management organisations to make decisions and take actions that increase robustness, redundancy, resourcefulness, and rapidity. Social resilience involves reducing negative consequences on communities due to the loss of critical services. Economic resilience involves reducing both direct and indirect economic losses caused by the occurred disruption. In this HDR thesis, we will mostly refer to technical resilience, in an acceptation that revolves around four fundamental, partially overlapping concepts [42, 40, 43, 38], defined as follows.

Firstly, *robustness* refers to the *preparedness* of the system to absorb and resist disruptions, thanks to the abundance of resources (*i.e., redundancy*) and the capability (*i.e., resourcefulness*) to identify problems, establish priorities, and mobilise resources when conditions exist that threaten to disrupt some element of the system. Robustness usually refers to a situation in which the system does not leave its stable state of operation despite the occurrence of disruptions, which should be therefore fully absorbed by the system without significantly affecting its performance. *Vulnerability* is a term strictly related to robustness. Specifically, vulnerability can be considered the antinome of robustness [44]. Vulnerability is defined as the expected damage (consequences C) that occurs in a system when a hazardous event A occurs and the system is exposed to it [45]. The triplet (A, C, U) is the definition of *risk* [45], with U representing the *uncertainty* of the occurrence of event A. Vulnerability assessment tries therefore to quantify the uncertainty U as well as the severity of consequences C, with associated probability P, given the knowledge K of the actual state of the system under analysis, *i.e.,* $V = (C, P, U, K|A)$ [44].

Secondly, the terms *rapidity*, *graceful extensibility* or *flexibility* can all refer to the capability of the system to stretch its existing resources to meet priorities and achieve goals in a timely manner to contain losses in the presence of disruptions. In particular, rapidity refers to the time passing between the stable state of operation and the new state of operation associated with a reduced level of service, while the extensibility capability of the system determines the amount of performance loss associated with the new degraded state of operation. The residual capacity associated with the degraded state is normally identified with the term of *survivability*.

Thirdly, *recovery* is a central aspect of resilience, corresponding to a system's capability to rebound to normal operation from the degraded state of equilibrium reached in the aftermath of a disruption. Recovery capabilities are largely dependent on the procedures, structures, and resources that have been developed before the occurrence of a disruption to handle expected or unexpected emergency situations. They are especially relevant when the event falls outside the scope of variations and disturbances that the system in question is known to be capable of handling. The traditional notion of recovery corresponds to the definition of ecological or static resilience, reflecting whether the system returns (or not) to the same state of function after some external perturbation or shock. A second, more recent notion of recovery corresponds instead to engineering or dynamic resilience, which refers to the rapidity of this recovery to a new level of operation [46]. It is worth highlighting that the new recovered state may not necessarily correspond to the initial normal state, as it

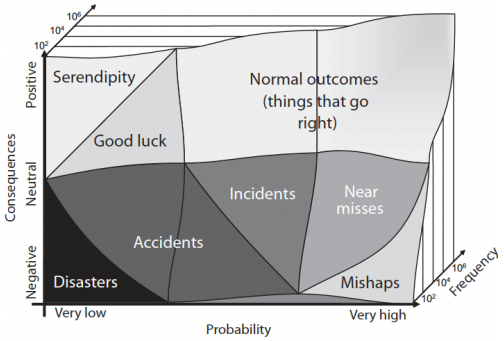
can be associated with lower or higher levels of performance with respect to the situation preceding the disruption.

Fourthly, *adaptability* refers to the capability of the system to sustain adaptation over its life cycle, by changing its internal organisation, structure and components with respect to predictable changes or sudden disruptions. This notion requires the system to be flexible in its architecture in order to sustain changes over long scales, which normally include multiple cycles of disruptions and failures.

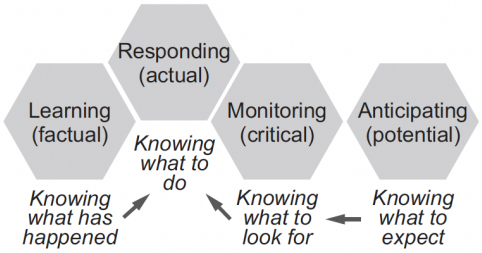
It is important to note that these concepts are strictly interrelated and that a resilient system should exhibit all these characteristics to varying degrees, depending on the specific risk situation the system is exposed to (see Fig. 1.4(a)). We will explicitly refer to these properties in this and the following chapters to clarify how the developed approaches and tools support specific components of transport resilience.

1.2.1 Main Approaches to Assessing Resilience: a High-level Classification

The prologue of the book on resilience engineering from Hollnagel *et al.* [13] provides a valuable conceptual framework to classify relevant approaches from the literature on the subject of resilience quantification and management.



(a) The relationship between probability of occurrence and consequences of outcomes in a system's life-cycle



(b) The four resilience cornerstones

Figure 1.4: The general framework for resilience engineering proposed by Hollnagel *et al.*, 2006.

Hollnagel *et al.* present resilience engineering as a field encompassing traditional safety (or risk) management. Resilience engineering focuses in fact on *both expected and unexpected conditions rather than just avoiding failures*, as normal outcomes are way more probable than extreme events, incidents and disasters (see Fig. 1.4(a)). According to the authors, it is in fact *necessary to look at success as well as at failures precisely in order to understand failures or why things can go wrong*. This represents a very interesting definition that clarifies how resilience engineering not only deals with safety but also with system efficiency, and that no fundamental conceptual differences exist between performance that leads to failures and performance that leads to successes. The framework also provides a clear description of what kind of operational solutions one should consider for improving system resilience. As graphically presented in Fig. 1.4(b), resilience engineering covers four main cornerstones, namely: *(i) knowing what to do*, *i.e.,* the set of prepared strategies to *stretch* and *recover* in the presence of *actual* sporadic or recurrent disruptions; *(ii) knowing what to look for*, *i.e.,* the capability to *monitor* the actual or potential *critical* (external or internal) threats a system is exposed to; *(iii) knowing what to expect*, *i.e.,* the set of features that help to *anticipate potential* future threats, changes and opportunities by means of *adaptation*

capabilities; *(iv) knowing what has happened*, *i.e.,* the ability to *learn* from past *factual* success or failure stories in the system's life-cycle.

Similar ideas were advanced by Bellini *et al.* in [14], where resilience-engineering is considered as composed of four key elements: *respond*, *anticipate*, *monitor*, and *learn*. The authors suggest that digital technologies can enhance the efficiency and effectiveness of resilience-building processes. For example, cloud computing, big data, and simulation tools can be used to *anticipate* potential risks and test emergency responses. IoT sensing and 5G technologies can improve urban *monitoring* and warning systems. Real-time data processing can contribute to more efficient *emergency responses*, while advanced data analysis techniques based on machine learning and optimisation can support *learning* and policy-making. The literature generally refers to this approach as *resilience analytics* [47, 48], which involves using data-driven processes to support resilience through descriptive, predictive, and prescriptive modelling.

In their recent survey [49], Serdar *et al.* analyse the abundant literature on resilience assessment for urban transport, identifying the main approaches adopted so far in the field. The authors identify five main categories of approaches: *big data methods and technologies*, *simulation methods*, *model optimisation*, *graph theory*, and *probability-based methods*.

Big data methods and technologies involve the use of large volumes of data from various sources, such as social media, mobile phone and GPS data, to analyse and monitor the performance of transportation networks in real-time. The data can be structured, semi-structured, or unstructured, and can be processed through different analytical tools and algorithms to detect abnormalities and reflect on past events. The authors consider big data methods suitable for assessing the resilience of transportation networks in terms of system performance, and they can provide valuable insights for smart city applications. However, the effectiveness and resource consumption of big data methods largely depend on using the correct indicators and data type, being often not capable of predicting future performance with respect to unprecedented events. Moreover, the authors insist on the absence of a good understanding of the actual, multi-modal and time-varying travel demand, which is a fundamental component that needs to be learned in order to anticipate and recover from disruptions.

Simulation methods are identified as another crucial category of approaches for assessing the resilience of transportation networks under fresh developments and unprecedented events. They involve creating models of transportation networks and testing them under different scenarios and traffic configurations. Simulation methods can consider different network components and their interactions, but they are resource-demanding, which makes it hard to scale up. They also require calibration and a deep understanding of the system to apply them correctly.

Model optimisation is considered suitable for post-disaster planning, preparedness, and recovery budget allocation. The approaches falling in this category involve developing mathematical models of transportation networks and optimising them to allocate resources efficiently. Model optimisation methods can be used to optimally schedule and prioritise the allocation of resources and evaluate the effectiveness of different mitigation plans. However, developing an efficient mathematical representation of the system requires a good understanding of both the system and the threats it faces. Methods in this category often combine with graph-based approaches for modelling the underlying transport network as well as simulation to compute generalised costs included in the objective functions or in the problem constraints.

Graph theory approaches to resilience engineering traditionally involve using network topology modelling and connectivity analysis to study and monitor the performance of transportation networks. Methods in this category generally adopt a simplistic representation of the transport system, which has however the important advantage of resource efficiency, thus being particularly suitable for analysing large networks. Graph theory methods do not naturally account for traffic dynamics, congestion formation, redistribution, and capacity

Table 1.3: Methods for resilience assessment, adapted and extended from Serdar *et al.*, 2022.

Assessment method	**Remarks**	**Related Chapters**
Big data, IoT, Machine Learning	**Suitability/Strength** • Suitable for smart city applications and resilience assessment through real-time system performance assessment, detecting the abnormalities, and reflecting on previous events. • Using correct indicators and the type of data (e.g., structured, semi-structured, or unstructured) play a central role in its effectiveness and resource consumption. **Limitations** • Under-exploited for understanding and modelling of multi-modal, time-varying travel demand. • Could be not suitable for predicting future performance toward unprecedented events.	Chapter 2, Chapter 3, Chapter 5
Simulation	**Suitability/Strength** • Suitable to assess fresh developments and unprecedented events, can consider different traffic compositions, network components, and their interactions. **Limitations** • Resource-demanding, which makes it hard to scale up. • Require calibration and deep understanding to apply it correctly.	Chapter 4, Chapter 5
Graph theory (complex networks, network topology)	**Suitability/Strength** • Suitable for large networks since it is simple and resource-efficient. • Can be used in conjunction with other methods. **Limitations** • Usually does not account for traffic performance, travel demand, and the dynamics of the urban transport system (traffic redistribution, congestion formation, *etc.*) • Basic static modelling of disruptions, mostly based on edge cut and nodes' removal. Does not typically take into account partial capacity reduction deriving from disruptions.	Chapter 3, Chapter 4, Chapter 5
Model optimization	**Suitability/Strength** • Suitable for post-disaster planning, preparedness, and recovery budget allocation. **Limitations** • Require good understanding of the system and threats to develop an efficient mathematical representation of the system.	Chapter 4
Probability-based methods	**Suitability/Strength** • Suitable for assessing system resilience in terms of its reliability. • Help to evaluate different arrangements and prioritizing severe disturbances mitigation plans. **Limitations** • Rely on the accuracy and representativeness of the statistical data used in the development of probability estimations.	Chapter 5

reduction due to accidents, which represents one of the most relevant limitations of these approaches, usually requiring them to be used in conjunction with other methods, such as traffic simulation and big data processing. The modelling of time-related dynamics can be addressed with more complex modelling approaches such as dynamic and multi-layer networks.

Probability-based methods involve assessing the resilience of transportation networks by modelling the probability of failures as well as the causal relationships among the system components, usually leveraging Bayesian models or data-derived statistical models of system performance. They help to evaluate different arrangements and prioritise severe disturbance mitigation plans. The accuracy of probability-based methods strongly depends on the representativeness of the statistical data used in the development of probability estimations.

Tab. 1.3, borrowed and partially adapted from [49], summarises the main conclusions from the authors and identifies the chapters of this HDR thesis that deal with each of the mentioned categories.

1.2.2 Vulnerability Analysis of Transport Networks

In the following, we propose a concise literature review, largely based on [50, 46, 49], about quantitative approaches and metrics for vulnerability analysis of urban transport networks, leaving the interested reader to the numerous surveys on the broader topic of network resilience engineering for more details [13, 38, 51, 47, 52].

In urban transportation, vulnerability refers to the degree of exposure of a system to hazards and losses due to the risks present in its operating environment [53]. Analysing the vulnerabilities of a system corresponds to the *knowing what to expect* dimension in the general framework proposed by Hollnagel *et al.* [13] (see Fig. 1.4(b)).

In [43], one of the first general-purpose methods to model and quantify vulnerabilities by measuring the impact of disruptions on a complex system was introduced. The framework also allows evaluating the positive effect of recovery strategies on the disrupted system (*i.e.,* the *knowing what to do* cornerstone). Specifically, the system under analysis is assumed to be stressed at time t_e by a disruptive event e_j that causes service loss to the system. As a result, the system shifts from the original state of regular operation s_0 to a disrupted stable state s_d at time t_d. Recovery actions are taken at time t_s bringing the system to a recovered state at time t_f. The transition sequence is graphically described in Fig. 1.5, borrowed from [43].

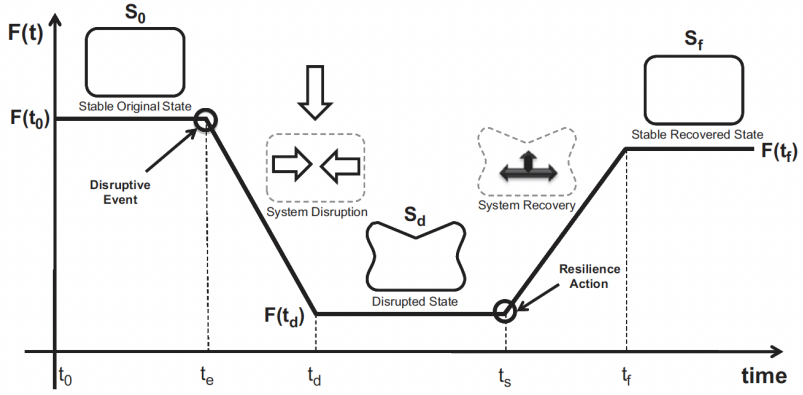


Figure 1.5: Transition among states of a disrupted system as proposed in Henry *et al.*, 2012.

The resilience $R(t_r)$ of the system at time $t_r \in (t_d, t_f)$ is thus defined as the ratio of recovery at time t_r to loss suffered by the system at time t_d:

$$\mathcal{R}(t_r) = \frac{Recovery(t_r)}{Loss(t_d)} \tag{1.1}$$

The main idea behind this generic definition of resilience is that if the recovery equals the loss, the system is entirely resilient. If no recovery action occurs, then the system has shown no resilience. The value of resilience can be computed based on a figure-of-merit $F(\cdot)$ representing a generic quantifiable indicator of whole system performance, such as

connectivity, throughput, reliability, *etc*[3]. Resilience can be thus computed at time t_r as:

$$\mathcal{R}(t_r|e_j) = \frac{F(t_r|e_j) - F(t_d|e_j)}{F(t_0) - F(t_d|e_j)}. \tag{1.2}$$

In the previous equation, if $F(t_r|e_j) = F(t_0)$, the value of resilience $\mathcal{R}(t_r|e_j)$ equals 1, meaning a fully restored system. In a similar approach, Bruneau *et al.* [42] propose to compute the area under the curve $1 - Q(t)$ to quantify resilience and compare different recovery strategies, specifically:

$$\mathcal{R}(t_r) = \int_{t_0}^{t_r} [1 - Q(t)]\, dt. \tag{1.3}$$

where $Q(t)$ represents the fraction of nominal performance exhibited by the system during disruption, *i.e.*, $Q(t) = \frac{F(t|e_j)}{F(t_0)}$, using the formalism from [43]. Sec. 4.1 will build upon this approach to propose a solution to simulate the impact of day-to-day perturbations on transport networks [55].

Similarly to Bruneau *et al.* (Eq. 1.3), Mattsson and Jenelius introduce in [46] the concept of conditional transport vulnerability, defined as the aggregate consequence of a disruption scenario. It is measured by the area between the dotted (blue) line of nominal behaviour and the relevant curve representing the reduced level of function, as shown in Fig. 1.6. The latter curves depend on *ex ante* mitigation or *ex post* adaptation actions. By knowing the fragility points of a transport system, adequate proactive actions can be envisioned, thus contributing to strengthening network resilience.

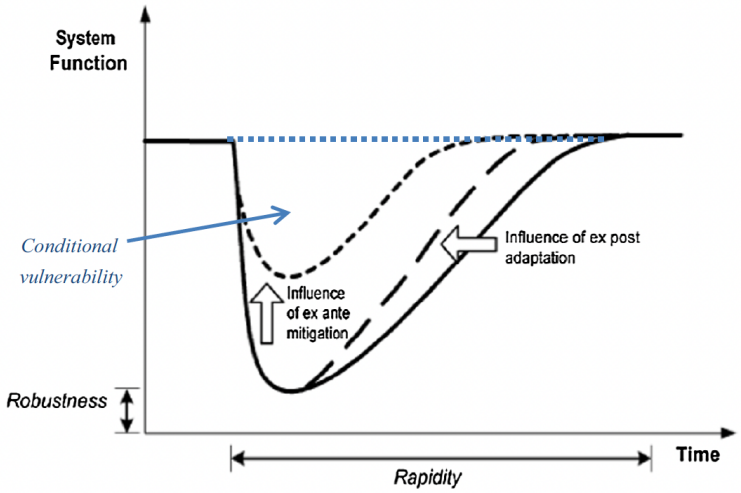


Figure 1.6: Resilience curve and conditional vulnerability as from Mattsson and Jenelius, 2015.

Based on these general concepts, multiple approaches and metrics have been proposed for analysing and quantifying transport network vulnerabilities. As proposed in [46], they can be categorised into two main groups: *topological* and *system-based*, detailed in the following.

Topological Vulnerability

In topological approaches, graph theory is used to model the transport network. The latter is represented as a graph, *i.e.*, $G = (V, E)$, with V denoting the set of nodes and E

[3] In our context related to transportation, candidate metrics could be traffic flow, the inverse of the total travel time, network efficiency [54], *etc.*, depending on the specific properties one would like to highlight in the definition of resilience.

the set of links connecting the nodes. When a transport network is composed of a single mode, simple graphs can represent it. However, when multiple transport modes (*e.g.,* bike, car, bus/tramway/subway/railway lines, *etc.*) and/or the entire set of Public Transport Network (PTN)/railway lines need to be represented at once, more complex representations, such as multi-layer network models, are necessary [56, 57]. In this case, a model such as the one introduced in Sec. 3.3 for multi-modal path reconstruction via TRAjectory inference at scale with mobile Network SIgnaling daTa (TRANSIT) is used. For a mono-modal/mono-line transport network, nodes represent stations in the case of PTN/railway, or road intersections in the case of the road infrastructure. Edges represent the connections between transit/railway stops or road segments between pairs of road intersections, respectively. Note that multiple formalisms exist, such as *L-space* and *P-space*, for representing transit networks, as widely discussed in [58]. In some works, such as [59], a dual network representation is used, with streets identified as nodes and intersections identified as links of unit length. Similar approaches are used in modelling air transport networks (which are not covered in the research activities described in this HDR thesis), with nodes representing airports and edges representing direct connections between airports.

Graphs are traditionally unweighted or weighted by means of attributes, such as road length, segment capacity, free-flow travel times, *etc.*, that are *static*, *i.e.,* do not change (or change very slowly) over time. In topological approaches, it is typical to use complex networks metrics to identify the most important, and therefore vulnerable, components of the transport network topology. Similarly to a reverse percolation process [60], disruptions are traditionally modelled via removing edges or nodes from the network topology according to random or targeted attack strategies [61]. The goal is to identify the specific set or the critical ratio of components that, once removed, would make the system incapable of performing its operations. In targeted attack strategies, nodes or edges are removed based on degree, betweenness or closeness centrality-based[4] ranking [61, 63, 64]. Indicators used to measure the impact of modelled disruptions include the size of the largest connected component, the average path length, the global clustering coefficient, global and local efficiency [64, 65, 66]. Topological approaches have been consistently applied to road, public transport and air transportation networks. These researches have usually indicated *small-world* properties[5] of road networks [61, 67, 38], which thus appear very robust against attacks based on random and degree strategies, while very vulnerable against attacks based on the betweenness strategy, *i.e.,* nodes and edges traversed by a high number of shortest paths should be significantly protected. Concerning transit, different works have proven the *scale-free* properties[6] of some real-world transport networks (*e.g.,* public transport of Bejing [68], public transport network of several European cities [69], subway system of Shenzen [70]), which indicates resistance to random failures but vulnerability to attacks on high-degree nodes; absence of either scale-free or small-world properties for some subway networks [71, 72], which indicates

[4] *Degree centrality* is a node-based centrality measure that counts the number of edges incident to a node. *Betweenness centrality*, on the other hand, is an edge or node-based centrality measure that quantifies the extent to which a node or edge lies on the shortest paths between other pairs of nodes in the graph. *Closeness centrality* is another node-based centrality measure that quantifies how quickly a node can reach other nodes in the graph. It is the inverse of the sum of the shortest path distances from a node to all other nodes in the graph. For formal definitions, the reader can refer to [62]. Betweenness centrality is widely discussed in Sec. 4.1, 4.2.2 and 5.4. Degree centrality is instead more extensively studied in Sec. 4.2.1.

[5] A small-world network is characterised by high values of global clustering coefficient, *i.e.,* neighbours of a node tend to be neighbours in turn, combined with a small average path length between pairs of nodes. Small-world networks are often observed in real-world networks, such as social networks, where there is a high degree of local clustering, but also a relatively short path length between any two individuals in the network.

[6] A scale-free network is characterised by a highly-skewed degree distribution, where a small number of nodes have a very high degree, while the vast majority of nodes have a much lower degree. Scale-free networks are often observed in real-world networks, such as the World Wide Web or biological networks, where a few highly connected nodes (known as *hubs*) play a disproportionately important role in the network's structure and function. Formally, a network is considered to be scale-free if its degree distribution follows a power law distribution, *i.e.,* the probability $P(k)$ of a node having degree k is proportional to $k^{-\gamma}$, where γ is a parameter that characterises the degree distribution of the network.

high fragility of such transport infrastructures. Scale-free properties have also been observed in air transportation networks [73].

System-based Vulnerability

Despite the interesting results, the main limitation of topology-based vulnerability approaches is the oversimplified representation of the transport system that completely neglects both travel demand and the dynamics of traffic operations (see Sec. 4.1). Resilience is in fact not purely a structural problem: for instance, road networks can easily become gridlocked in presence of natural disasters that do not impact at all the physical infrastructure (roads, bridges, and railways), but that totally shatter the typical mobility demand patterns or the supply-side of the transport system. In other words, there is strong evidence that one must monitor and control the dynamic states of the system in order to properly address the resilience of networked CI [38].

To capture the dynamic nature of critical infrastructure systems, time-varying graphs have been recently utilised to incorporate time-dependent interdependences into resilience metrics: studies by Adjetey-Bahun *et al.* [74] and Lu [75] have compared the performance of static and dynamic resilience metrics concluding that, under normal traffic conditions, both metrics are equally effective. However, in the event of a disruption, interdependences and passenger flows in the network render the static indicator less efficient and highlight the need to consider the time dimension in vulnerability analysis. These studies represent the basis of the modelling approach described in Sec. 4.2.

The system-based category of vulnerability analysis approaches involves the representation of transport systems through demand and supply models, utilising micro-, macro- and mesoscopic traffic simulators to generate realistic time-varying variables such as travel times, flow, and cost variables. These variables are then used to compute vulnerability indicators. Additionally, network equilibrium is used to model route choice, taking into account congestion and delays. Approaches in this category, mostly related to road network vulnerability, consider disruption modelling via node or edge removal similar to topological-based methodologies. However, they differ in that they exploit the aforementioned demand-supply models and simulation tools to generate realistic system dynamics and identify the most important links or nodes with respect to such dynamics. Critical nodes or edges are defined as those that, if removed, would have the most significant impact on either increasing the shortest path duration or reducing the maximum flow between a given set of origin-destination pairs. More advanced approaches consider the transport system's performance with respect to fluctuations [76] or significant variations [77] of travel demand or link capacity. Performance is evaluated based on indicators such as travel time reliability (*i.e.,* the probability that a trip can be completed within a specified time interval) and capacity reliability (*i.e.,* the probability that a network can accommodate a specified level of travel demand).

Other approaches in this category propose different vulnerability indices. Researchers have used the increase in the user-equilibrium (total) travel time [78, 79], the reduction of the demand weighted global efficiency [80], the time to link re-opening [81] and the amount of unsatisfied demand [78] to measure the decrease of service caused by the loss of a road link. Similarly to what we propose in Sec. 4.1, some isolated works have also considered more flexible modelling of disruptions with partial link capacity reduction modelling [82]; aggregate indices of vulnerability to characterise the exposure to risks of an area or of the entire transport network [83, 84], (see Sec. 4.2.1); indicators of local vulnerability only accounting for travel time increase for trips originating in a given region, when links in the network are randomly closed [85]; indicators of global vulnerability derived from local events by computing the total travel time increase observed when all links in a region are closed [83]. The last approach is particularly relevant to model area-covering disruptions deriving from snowstorms, floods and earthquakes.

Vulnerability has been also measured in terms of accessibility reduction or remoteness

increase [86, 87], by also taking into account spatial data on population [88] in the definition of accessibility. Some authors have studied the correlation between link importance metrics and static topological properties of the link, (*e.g.,* free flow speed, capacity, length, *etc.*) [85, 89]. According to probabilistic-based methods and historical data, some researchers have modelled the development of traffic flows after network disruption based on past experience and beliefs about future conditions [90]. Other approaches adopt mathematical modelling and optimisation techniques to identify worst-case scenarios and best responses to the latter. Solutions based on attacker/defender two-player games [91] and bi-level optimisation [92] have also been proposed.

System-based approaches for vulnerability assessment have been considered for rail and public transport networks. Subway and rail networks have been proven to be more sensitive to disruptions than road networks because of the lack of excess capacity and the limited amount of alternatives to redirect trains in case of disruptions. These features often lead to total loss of service and high delays on subsequent trains. For instance, Rodríguez-Núñez and García-Palomares [93], have evaluated among others, the importance of links in a subway system and quantified the amount of unsatisfied demand in terms of trips following the closure of the most important ones.

To account for the interactions between supply and demand and the accumulated effect of disruption on PTN, Cats and Jenelius [94] have introduced a novel measure of betweenness centrality that takes into account the expected number of trips and passengers between pairs of stops over time periods. This dynamic and extended definition of betweenness centrality is identified as a better indicator to evaluate the impact of disruptions on passenger welfare than the static betweenness centrality, which is based solely on network topology. In a follow-up work [95], the same authors investigated solutions for reducing the vulnerability of the public transport network by increasing the capacity on lines that can serve as alternatives when critical links are disrupted, and by identifying the lines where capacity increases are the most effective.

Peterson and Church [96] developed a framework for modelling rail freight transport vulnerability with similar approaches to those proposed in [78, 79] for road networks, based on the notion of link importance. A statistics-based framework is proposed by Hong *et al.* [97] to study the vulnerability of the Chinese railway system to floods and the effectiveness of alternative mitigation strategies. Both studies highlight the relevance of betweenness centrality as a means for selecting candidate links for vulnerability assessment and resilience enhancement. Solutions for air transport based on the quantification of delayed flights versus on-time flights during disruptions are considered in [98].

Overall, vulnerability analysis using system-based solutions is generally less straightforward when compared to topological approaches, as it requires *(i)* more computational resources to simulate system dynamics, *(ii)* massive data to acquire knowledge on the transportation system's dynamic states. However, it is well known that system-based solutions typically yield more realistic and reliable results because they incorporate both supply and demand-side modelling in dynamic settings, and because they allow for a more accurate and wider-spectrum representation of the consequences of disruptive events. These advantages well-justify the growing interest of the research community towards these approaches, which tend however to be increasingly used in conjunction with topological and big data processing techniques for efficiency purposes. These ideas have been at the very core of the research described in the next chapters, which propose solutions to evolve vulnerability analysis (Chapter 4) and monitoring towards large-scale and real-time settings (Chapter 5).

In any case, both categories of approaches have traditionally ignored the complex multi-modal nature of the transport system. Topological approaches have the merit to be more easily adaptable to multi-layer networks that can be used to model multi-modal transport networks. Some recent works have thus considered the problem of performing vulnerability analysis on multi-modal transport networks [57, 99], using centrality metrics, extended to multi-layer networks, to perform structural static analysis of multi-modal transport networks

exposed to disruptions. Despite not explicitly addressing the problem of disruption analysis, Bellocchi, Latora and Geroliminis [100] have very recently proposed a system-based measure, called dynamical efficiency, to evaluate the performance of a multi-modal time-varying transportation network. The metric considers both the topology of the network and the temporal changes in traffic flow to detect critical zones for traffic congestion and bottlenecks in a transportation system. Dynamical efficiency for a given layer (*i.e.,* a single transport mode) is defined by considering the average ratio between the minimum travel time (*i.e.,* free-flow travel time) of a path between any two nodes in the network to the actual travel time of that path over a time period. The metric is extended to multi-modal settings by considering shortest $\mathcal{H}$ *paths*, introduced by the authors to reduce the complexity of multi-modal shortest path computation. Based on this notion, additional derived metrics, such as the layer performance gain and station centrality, are proposed to quantify the relevance of an entire transport mode or of specific stations of a multi-modal network. The effectiveness of this approach has been verified on a real-world multi-layered transportation system, from the city of Shenzhen, China. The introduced metrics appear to allow for the identification of spatio-temporal congestion patterns, the quantification of the expected usage of inter-modal junctions between two different transportation means, as well as the assessment of equivalent transportation alternatives. These ideas, being already partially addressed in the modelling framework of the PROMENADE platform described in Chapter 5, are also at the centre of the longer-term perspectives which will focus on extending the notion of resilience assessment and enhancement to a multi-domain context (Chapter 6).

Chapter 2

Mobile Traffic Profiling and Travel Demand Estimation

Over the last two decades, global mobile traffic has surged to unprecedented levels, with a compound annual growth rate higher than that observed for Internet traffic during the surge of the World Wide Web at the turn of the millennium. Current trends indicate this growth persists, primarily driven by the global deployment of 5G technology. Within the last two years alone, mobile network traffic has doubled. This is largely a result of increased video content consumption and the rise in mobile subscriptions [101]. Projections indicate that by 2028, 5G subscriptions will reach 5 billion, accounting for 55% of all mobile subscriptions.

The surge in mobile demand poses substantial challenges for mobile network operators, necessitating effective management of the initiation, modification, release, and relocation of network resources [102]. To address these challenges, there is an urgent need for data-driven algorithmic approaches that can analyse and predict mobile usage, specifically the demand for communication services from users. Simultaneously, the abundance of probe data collected by network operators, coupled with the richness of the associated spatio-temporal information, as discussed in Chapter 1, can significantly enhance our understanding of both communication and transportation demand.

This chapter explores the utilisation of mobile phone data from two closely related perspectives: firstly, understanding cellular service demand through analytical tools to effectively manage network resources, and secondly, investigating the potential of mobile phone data to enhance our understanding of travel demand. Due to the evolution of data collection technologies from mobile network providers, the research activities associated with the first perspective leveraged CDR data, while the second perspective focused on the use of new-generation passively collected mobile phone data, *i.e.,* NSD.

During my postdoctoral research at the CITI-lab of INSA-Lyon, I developed unsupervised strategies to classify usages of aggregated data related to historical mobile phone traffic demand, specifically CDR. These insights are concisely showcased in the initial part of this chapter (Sec. 2.1). My investigations have shed light on the intricacies of mobile traffic demand and introduced methods for the sustainable roll-out of 5G telecommunication infrastructures. Such endeavours have formed the foundation for my subsequent research, predominantly centred on resilience modelling for urban transport networks.

The second part of this chapter pivots to the promising potential of mobile phone data, particularly NSD, in the field of transportation studies. The research discussed in Sec. 2.2, corresponding to the initial activities at the LICIT-ECO7 laboratory, had as objective to explore NSD to supplement traditional household travel surveys and provide a description of the expected mobility demand. Specifically, we studied the applicability of signalling data in generating OD matrices. This data was sourced from millions of anonymous mobile

phone users in the Rhône-Alpes region, France. Notably, unlike CDR data, signalling data provides a broader spectrum of network-based records, ensuring superior spatio-temporal granularity. The results presented in the second part of this chapter (Sec. 2.2) have been the outcome of research collaborations with Orange Innovation and tutoring activity of one PhD student (*i.e.,* Mariem Fekih). The data used in the context of OD flow estimation were made available by Orange Innovation in the context of Mariem Fekih's PhD thesis.

This chapter includes content from the following papers:

1. A. Furno, D. Naboulsi, R. Stanica, and M. Fiore, "Mobile demand profiling for cellular cognitive networking", IEEE Transactions on Mobile Computing, vol. 16, no. 3, pp. 772–786, 2016.
2. A. Furno, M. Fiore, R. Stanica, C. Ziemlicki, and Z. Smoreda, "A tale of ten cities: Characterizing signatures of mobile traffic in urban areas", IEEE Transactions on Mobile Computing, vol. 16, no. 10, pp. 2682–2696, 2016.
3. A. Furno, M. Fiore, and R. Stanica, "Joint spatial and temporal classification of mobile traffic demands", in IEEE INFOCOM 2017-IEEE Conference on Computer Communications. IEEE, 2017, pp. 1–9.
4. M. Fekih, T. Bellemans, Z. Smoreda, P. Bonnel, A. Furno, and S. Galland, "A data-driven approach for origin–destination matrix construction from cellular network signalling data: a case study of Lyon region (France)", Transportation, vol. 48, pp. 1671–1702, 2021.
5. M. Fekih, L. Bonnetain, A. Furno, P. Bonnel, Z. Smoreda, S. Galland, and T. Bellemans, "Potential of cellular signaling data for time-of-day estimation and spatial classification of travel demand: a large-scale comparative study with travel survey and land use data", Transportation Letters, vol. 14, no. 7, pp. 787–805, 2022.

2.1 Mobile Traffic Demand Profiling with Call Detail Records

This section describes the research aimed to improve the current understanding of the consumption of mobile network resources in space and time, and to show and describe the existence of correlations between mobile telephony data, the characteristics of human mobility and the socio-economic functions of urban environments. Orange, one of the most popular mobile operators in France, shared several months' worth of anonymised CDR for several French, Ivorian and Senegalese cities [103, 104]. Other data were made available for several Italian cities by Telecom Italia Mobile (TIM), one of Italy's leading operators in the sector. Such data were provided in the context of the TIM Big Data Challenge [105], to which our team has participated, being awarded as one of the top projects. The CDR used in this research described the total volume of mobile traffic in terms of the number of voice calls and messages exchanged between any two of the operator's base stations, aggregated on an hourly basis. It is interesting to note that this data is spatio-temporal: the information is provided with the latitude and longitude of the base station where the event was recorded, together with a timestamp indicating when a certain amount of traffic was observed.

2.1.1 Time Profiling of Mobile Traffic Demand

Based on previous work from the Urbanet team [7], we proposed in [6] a framework for the automated discovery of mobile demand profiles in large-scale cellular networks. The framework allows processing CDR for constructing sensible categories of the demand that

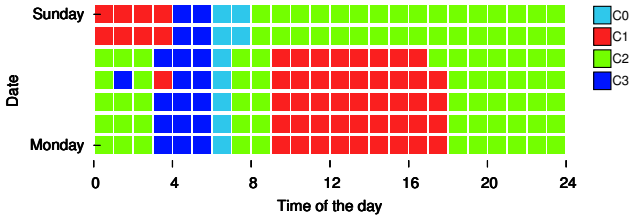
are associated to macroscopic spatio-temporal routines of the user population. As an interesting by-product, the framework can identify unusual behaviours in the demand, caused by events (*e.g.,* sudden traffic surges, reduced overall communication activity, etc.) biasing the customary dynamics of a significant subscriber fraction. As further detailed next, such events are typically related to public celebrations, sport events, concerts, holiday breaks or strikes, showing the sensitivity of CDR to follow and, in some cases even anticipate, the unfolding of urban events that move masses of people within or outside the city.

In a first phase, named *the training stage*, the framework leverages basic data preprocessing techniques (*i.e.,* data filtering and median aggregation) to extract hourly snapshots of *typical* user-generated mobile traffic over a city. Custom metrics of mobile traffic snapshot distance are employed to compute the similarity between pairs of snapshots by considering the geographical characteristics of the city. The framework makes use of hierarchical clustering with the average linking criterion to identify classes of related traffic snapshots. This hierarchical clustering outputs a whole family of solutions that can be represented as a dendrogram: it thus returns a richer information than a single-cluster set solution, as in the case of, *e.g., k*-means. However, this also implies that some criteria must be adopted to select the best clustering among all those in the family. To that purpose, a combination of well-known cluster analysis stopping rules determines the best number of clusters to retain. The approach is fully detailed in [6]. The training phase generates classes or schemes of typical network usages that can be used by a network operator to drive resource adaptation – via virtualisation solutions - in the mobile network. Moreover, such temporal classes have proven to possess a more general and social value, by showing insights on the regularities associated to human socio-economic activities, mobility patterns, and the different circadian rhythms that characterise the pulse of a city or a nation.

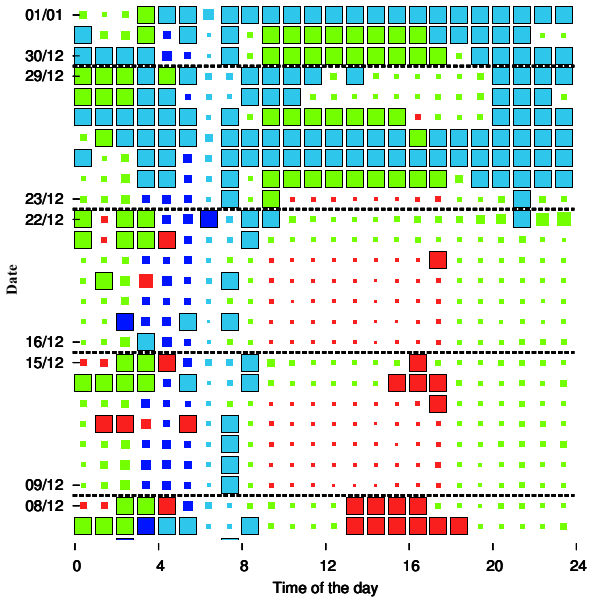
In its second phase, named *the classification stage*, the framework uses the previously identified schemes to classify the actual hourly CDR snapshots by relying on a modified implementation of the K-means clustering algorithm. Such step allows checking whether mobile traffic at given time belongs to the expected class and pinpointing anomalous situations otherwise.

The framework has been extensively evaluated over multiple CDR datasets collected from different world cities, including the data provided from Orange on Ivory Coast, Senegal and France, as well as the CDR data available for Italian cities from TIM. A few results are reported, as an example, in Fig. 2.1, for the case of a typical week of traffic in the city of Milan, Italy.

Fig. 2.1(a) graphically reports the network profiles identified by the framework during the training phase (*i.e.,* clustering of the hourly snapshots from the median week). Such classes can be naturally mapped to activities taking place in dense urban areas during office time or weekend night-life (C1), weekend-residential usages (C2), early morning behaviours typically related to commuting (C0) and deep night with reduced communication activity (C3). Fig. 2.1(b) shows instead the results of the online usage of the framework (*i.e.,* the classification stage). For each snapshot, the plot shows with a different colour the class it belongs to. Additionally, the plot includes two graphical representations for detecting outlying behaviours. First, snapshots that fall in a different category from the expected one are represented with big squares, surrounded by a black border (e.g., December 14^{th} between 15:00 and 18:00. Second, snapshots classified in the expected cluster (from Fig. 2.1(a)) are represented with squares whose size is proportional to the normalised average distance from the expected class (*i.e.,* the smaller the size, the more correct the classification). In the figure, most of the snapshots related to Christmas Day and other public holidays are classified as outliers in the C0 cluster, instead of C1. During such holidays, traffic activity is quite low in the city centre, with respect to typical working days, and mostly concentrated in transport and residential areas. This is explained by the fact that most of the business activities are closed and the presence of people in the urban area of Milan is significantly reduced with respect to typical days.



(a) Reference profiles obtained via clustering of median traffic 1-hour snapshots (on a weekly basis) and traffic distribution similarity. Each square represents one snapshot, and colours map to categories



(b) Online usage of the framework and detection of outliers

Figure 2.1: Mobile traffic profile categories for the CDR of the city of Milan.

Some other snapshots, which are very close in time to public holidays (e.g., Tuesday, December 24$^{\text{th}}$ or Monday, December 30$^{\text{th}}$, 9:00 to 18:00), are classified in the C2 cluster instead of C1. They present a higher activity in the city centre and a quite regular traffic distribution in all the other zones and therefore clustered together with evening hours or weekends.

The framework is also capable of detecting special events that occurred in Milan in December 2013. For example, on December 7$^{\text{th}}$ afternoon, the "La Scala" theatre season opening traditionally takes place in the city centre of Milan. The same day is also the feast day of Saint Ambrose, the patron saint of Milan, with several public and religious events celebrated in the city centre. Similarly, on Sunday, December 22$^{\text{nd}}$ at 20:45, one of the most important football matches of the 2014-2015 season, *i.e.,* Inter-Milan, was played in San Siro stadium, registering a record presence of 79,311 spectators.

It is worth highlighting that the described framework has been used as an input to subsequent research on 5G cognitive networking, specifically, on data-driven dynamic 5G-networks optimisation, in the context of a joint collaboration with CNAM Paris and the University of Milan [106]. Moreover, it has marked the initiation of a long-lasting collaboration with Orange France, which has proven crucial for the research results reported in Chapter 3.

2.1.2 Spatial Profiling of Mobile Traffic Demand

The research experience outlined in the previous section has been pivotal in consolidating my interest in large-scale data processing and sparked my curiosity to investigate further mobile phone data towards automatic detection of recurrent patterns and anomalies of human presence and activities in urban environments.

It is well known that individuals exhibit highly repetitive yet distinctive mobility and activity patterns influenced by factors such as family life, work obligations, hobbies, occupations, and personal habits, as well as the availability of infrastructures, services, and amenities. This naturally applies to mobile network subscribers and their communication patterns as well. The uniform yet varied behaviour of mobile users results in heterogeneity in subscribers' profiles [107], periodic temporal demand [7], load fluctuations during large-scale social events [108], and geographic diversity in mobile communications [109].

The next stage of my postdoctoral research was thus focused around these research questions: can CDR data be leveraged to automatically infer the multiple *urban fabrics* of a city? Is it possible to observe the same patterns from CDR in multiple cities or even countries? Urban fabrics relate to the combination of infrastructures, such as transportation and telecommunication systems, sports centres, educational institutions, and healthcare facilities, and land use, including residential, industrial, touristic and commercial areas, that characterise different zones within a metropolitan area. Previous research has confirmed significant correlations between mobile demand and city cartography, including the diversity of mobile activity within urban areas [110], high similarity in the temporal dynamics of traffic in residential zones [111], as well as the existence of relevant shifts in communication load peaks between different areas of the city throughout the day and during weekday-to-weekend transitions [112]. Furthermore, during the time the activities described in this section were undertaken, the use of mobile phone data, and particularly CDR, was gaining momentum as a means of validating theories on creating liveable cities in urban planning [113].

With these findings in mind and the experience of my latest postdoctoral research, I decided to further explore the spatial heterogeneity of mobile communication activities and investigate unsupervised solutions to automatically identify relevant partitions of large metropolitan areas from CDR. The central hypothesis was that the identified zones, distinguished by homogeneous patterns of mobile traffic usage, would exhibit equivalent or highly similar urban fabrics. Another challenge consisted in defining a novel quantitative description of urban fabrics based on the identified regularities (*i.e.,* a *signature*) of the observed mobile traffic demand in the associated zones. Signatures could be helpful to unambiguously and informatively describe the socio-economic activities in the identified areas, even in the presence of mixed usages.

This research resulted in the development of a second framework built upon the concept of mobile phone traffic signatures for the automatic identification of urban fabrics from CDR. The framework is succinctly outlined below, along with some noteworthy outcomes. The interested reader can refer to [9] for complete details.

The framework assumes the availability of a generic dataset $\mathcal{D}$ describing the communication activity of a mobile subscriber population during a set of days $\boldsymbol{\delta} = \{d\}$. For each day, the mobile demand is assumed to be known as the aggregate volume of the traffic generated by all users in the same area during a given time interval; the size of the area and duration of the interval determines the spatial and temporal granularity of the dataset, respectively. These assumptions are typically satisfied when working with traditional CDR data and their more recent extensions (*e.g.,* NSD). We name *unit area* the spatial aggregation level: the whole geographic region under consideration $\mathbf{a} = \{a\}$ is thus divided[1] into unit areas a. The time granularity is instead characterised by the duration of a *time slot*, *i.e.,* the interval during which user activity is aggregated in each unit area. Each day $d \in \boldsymbol{\delta}$ is thus split into

[1]The definition of unit area is general and can accommodate any tessellation of space. Unit areas can map to, e.g., cell sector boundaries, coverage zones of base stations, Voronoi cells, or elements of a grid.

a set $\mathbf{t} = \{t\}$ of time slots t. Overall, $\mathcal{D} = \{v_a(d,t)\}$, where every element $v_a(d,t)$ describes the total mobile communication activity observed within each unit area a at time slot t of day d. From such data, the framework includes two main stages aimed to: *(i)* building a representative profile of the typical activity taking place in each unit area of the region of interest; *(ii)* clustering unit areas with similar profiles for urban fabric detection.

Concerning the first phase, the raw traffic volumes of each unit area a are transformed into the Median Week Signature (MWS) of the unit area. For the formal definition of the MWS of the generic unit area a, let us first introduce the notion of weekly support $\boldsymbol{\delta}$, corresponding to the time span of one week, *i.e.,* $\boldsymbol{\delta} = \{\text{MON,TUE,WED,THU,FRI,SAT,SUN}\}$. Let us also denote as $\boldsymbol{\delta}^{\delta} \subset \boldsymbol{\delta}$ the set of days in the dataset $\mathcal{D}$ that correspond to the day of the week δ, with $\bigcup_{\delta \in \boldsymbol{\delta}} \boldsymbol{\delta}^{\delta} = \boldsymbol{\delta}$. For instance, $\boldsymbol{\delta}^{\text{MON}}$ groups all Mondays in the dataset. Then, the generic element in the MWS of unit area a is defined as:

$$s_a(\delta,t) = \mu_{1/2}\left(\left\{v_a(d,t) \mid d \in \boldsymbol{\delta}^{\delta}\right\}\right), \quad \forall a \in \mathbf{a}, \tag{2.1}$$

for time slots t during day of the week δ, and where $\mu_{1/2}(\cdot)$ represents the median of the set within parenthesis. The complete MWS is then defined as the concatenation of time-ordered samples on the temporal support $\boldsymbol{\delta}$ of area a, or, formally:

$$\mathbf{s}_a = ||_{\delta \in \boldsymbol{\delta}} \left(||_{t \in \mathbf{t}} s_a(\delta,t) \right), \quad \forall a \in \mathbf{a}. \tag{2.2}$$

where the $||$ operator indicates the time-ordered concatenation of all elements in a set.

The second stage of the framework builds on the clustering approach described in Sec. 2.1.1 using the same linkage clustering algorithm with the average distance criterion. A distance metric based on the Pearson correlation coefficient is used to compute the dissimilarity between pairs of MWS related to different unit areas, *i.e.*:

$$\Delta_{a,a'} = 1 - C_{a,a'}, \quad \forall a, a' \in \mathbf{a}. \tag{2.3}$$

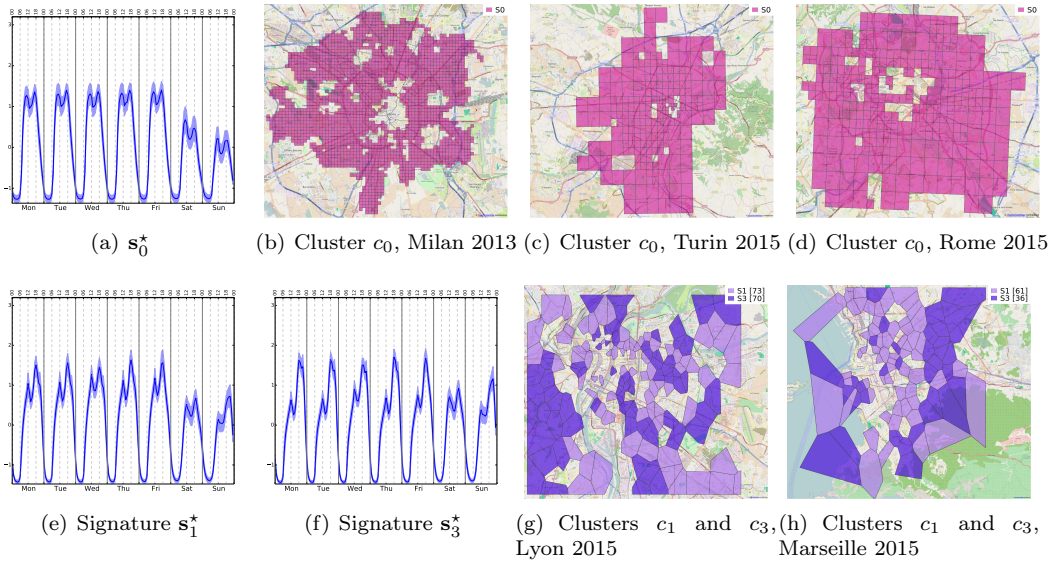
with $C_{a,a'}$ corresponding to the Pearson correlation coefficient of $\mathbf{s}_a$ and $\mathbf{s}'_a$. The skewness of the cluster sizes is evaluated at the different levels of the dendrogram built by the hierarchical clustering to select the number of clusters to retain: selecting the level with minimum skewness allows grouping unit area signatures into classes of relatively comparable sizes.

The assessment of the framework has been carried out on a comprehensive collection of datasets, including the CDR supplied by Orange France for six French cities and by TIM for four Italian cities over several months and years. This represents a much larger dataset than those ever used in previous studies, enabling us to generalise our findings and examine similarities and differences across a significant number of diverse cities. Moreover, the comparative evaluation reported in [114] has shown that our solution outperforms the related work [115, 116, 117], by identifying mobile demand profiles that better agree with reference data.

In Fig. 2.2, we report three major classes (c_0, c_1 and c_3) identified by our framework from the CDR records of the ten analysed cities. The characteristic signature $\mathbf{s}_0^{\star}$ of class c_0[2] is portrayed in Fig. 2.2(a). The comparison with Points of Interest (PoIs) and land use data from OpenStreetMap (OSM)[3] highlights the absence of any noticeable infrastructure and a dominant residential land use in the areas associated to cluster c_0, only related to Italian cities. This is outlined, in Fig. 2.2(b), Fig. 2.2(c) and Fig. 2.2(d), which show the extent of unit areas in Milan, Turin and Rome whose signatures are in class c_0. The corresponding regions include suburban and mainly residential zones of these cities and exclude city centres and well-known PoIs of the cities. Class c_0 can be thus associated with *residential urban fabrics in Italy*, often denoted by a mixture of private housing and small business activity.

[2]The characteristic signature of a cluster is computed as the average of all MWS signatures of unit areas belonging to the given cluster.

[3]https://www.openstreetmap.org



(a) $\mathbf{s}_0^\star$ (b) Cluster c_0, Milan 2013 (c) Cluster c_0, Turin 2015 (d) Cluster c_0, Rome 2015

(e) Signature $\mathbf{s}_1^\star$ (f) Signature $\mathbf{s}_3^\star$ (g) Clusters c_1 and c_3, Lyon 2015 (h) Clusters c_1 and c_3, Marseille 2015

Figure 2.2: Residential urban fabrics. Characteristic signatures (with standard deviation) and maps of related unit areas in representative city scenarios.

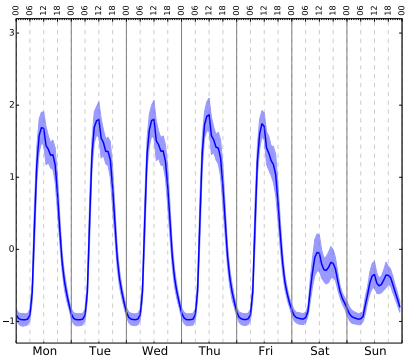
The characteristic signature s_0^* in Fig. 2.2(a) exhibits two comparable traffic peaks, at 11:00 and 17:00, repeating on all working days. The mobile activity in most urban areas in Italy is comparatively reduced during the weekend (but higher on Saturdays) with a morning peak higher than the afternoon one, which is also shifted towards later hours.

A similar discussion holds in the case of signatures classes c_1 and c_3, this time related to French cities only. These classes designate *residential urban fabrics in France*, as exemplified by the geographic coverage of the associated unit areas in Lyon and Marseille, shown in Fig. 2.2(g) and Fig. 2.2(h), respectively. Concerning the characteristic signatures $\mathbf{s}_1^\star$ and $\mathbf{s}_3^\star$, in Fig. 2.2(e) and Fig. 2.2(f), respectively, relevant similarities emerge in the semantics of the two signatures. Both feature two traffic peaks, the afternoon one standing over the morning one; the activity during weekends is comparable in the two cases, just scaled up in $\mathbf{s}_3^\star$. The main difference between $\mathbf{s}_1^\star$ and $\mathbf{s}_3^\star$ appears thus to be the afternoon-to-morning peak ratio, higher in the latter. We conclude that both c_1 and c_3 are representative of residential and small business areas in France, although c_3 is associated with a higher concentration of residential land use than c_1: indeed, the darker unit areas in Fig. 2.2(g) and Fig. 2.2(h), mapping to c_3, are more present in the urban outskirts and less so in city centres.

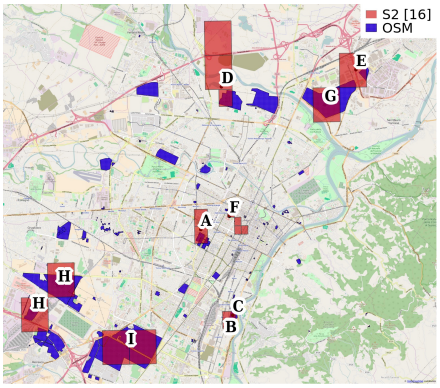
Comparing c_0, c_1, and c_3 is also intriguing. The disparities between the baseline profiles of mobile traffic demand in Italy and France are remarkable. The activity peaks are uneven and shifted by approximately one hour in France, and the ratio is even inverted during weekends. This diversity can be attributed to diverse routines in the two countries, raising interesting sociological questions.

Fig. 2.3 presents the characteristic signatures and maps associated with another set of major clusters identified by our framework, *i.e.*, c_2, c_5 and c_7. Interestingly, this time the clusters are not separated between nations, with unit areas from both France and Italy appearing in each of them. We relied again on OSM data as a source of information to understand the kind of urban fabrics associated with the detected clusters. When superposing the urban surface covered by unit areas associated with c_2 to OSM data, we notice a good match with locations mostly related to office-hour work activities, highlighted with blue-coloured polygons in the pictures[4]. Maps of exemplar case studies are provided in Fig. 2.3(b), Fig. 2.3(d), and Fig. 2.3(f) for Turin, Milan, and Paris, respectively. The anal-

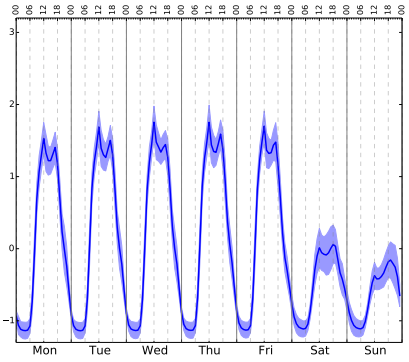
[4]Relevant PoIs are also marked with letters in the figures. The interested reader can refer to [9] for an exhaustive description of this PoIs.



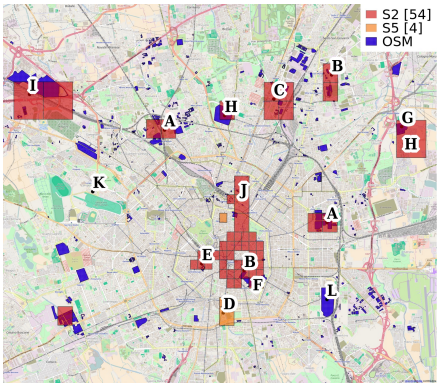
(a) Signature $\mathbf{s}_2^\star$



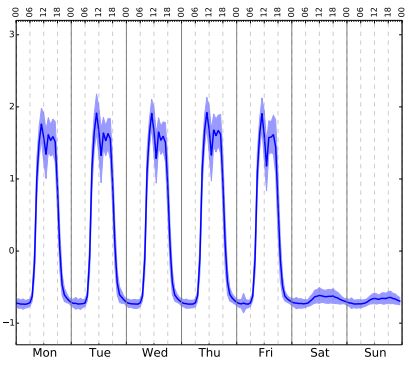
(b) Cluster c_2, Turin 2015



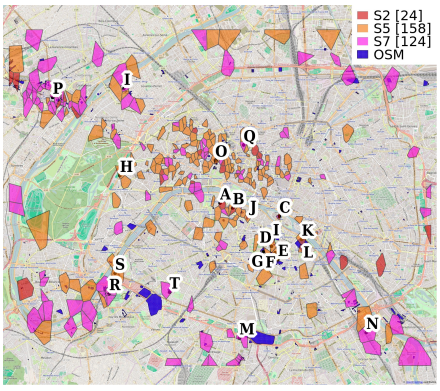
(c) Signature $\mathbf{s}_5^\star$



(d) Clusters c_2 and c_5, Milan 2015



(e) Signature $\mathbf{s}_7^\star$

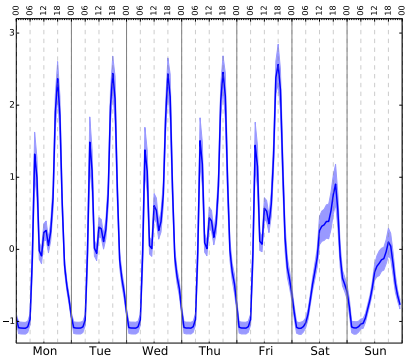


(f) Clusters c_2, c_5 and c_7, Paris 2015

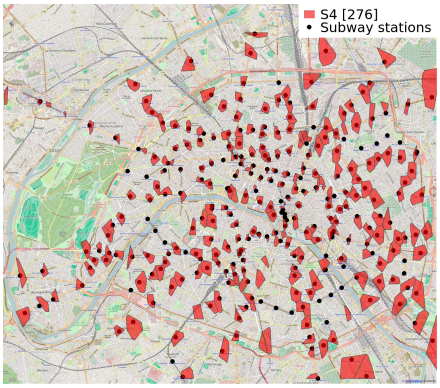
Figure 2.3: Office fabric signatures $\mathbf{s}_2^\star$, $\mathbf{s}_5^\star$ and $\mathbf{s}_7^\star$ and maps of the related unit areas in Italian and French cities, with OpenStreetMap data.

ysis of the related signatures (s_2^*, s_5^* and s_6^*) indicates a fairly constant and comparatively higher activity during office hours, especially during morning time, and drastically reduced during evening time, early morning and weekends, when a very small fraction of offices is open. Based on these considerations, it appears reasonable to consider these clusters as representative of *office urban fabrics*, *i.e.,* urban areas featuring socio-economic activities related to development, commercialisation and fruition of services and goods, with a typical European working time during weekdays, 9:00 - 18:00.

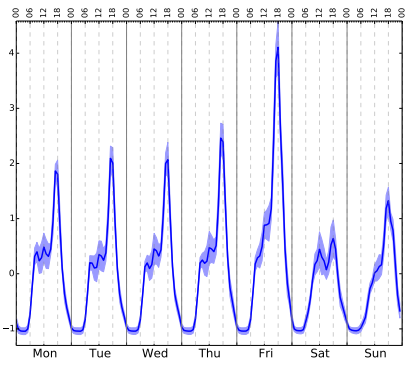
Finally, in Fig. 2.4, we report clusters that have shown a strong connection to *transportation urban fabrics*. This is, for instance, the case of cluster c_4 (Fig. 2.4(b)): this class



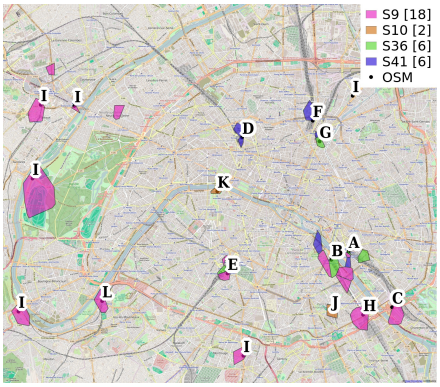
(a) Signature $\mathbf{s}_4^\star$



(b) Cluster c_4, Paris 2015



(c) Signature $\mathbf{s}_{36}^\star$



(d) Clusters c_9, c_{10}, other classes, Paris 2015

Figure 2.4: Transportation fabric signatures $\mathbf{s}_9^\star$, $\mathbf{s}_{10}^\star$ and $\mathbf{s}_{36}^\star$, and their maps for French and Italian cities, with OpenStreetMap data.

regroups unit areas with a very peculiar signature with two major peaks (Fig. 2.4(a)), clearly related to commuting, and a minor peak at noon, probably related to lunch activities around offices and schools. The analysis of PoIs retrieved from OSM indicated a striking match with subway stations of the city of Paris, reported in the figure as black dots. Similarly, the other clusters reported in Fig. 2.4(d) have an almost perfect match with train stations and appear associated with long-range commuting behaviours.

The aforementioned approach has been further refined, at the beginning of my activity in the LICIT-ECO7 research laboratory, to achieve finer spatial granularity by means of an original approach based on multi-source data, by fusing CDR and GPS taxi data. In [118], we proposed an original stage-based fusion technique, in which the clustering approach proposed for mobile phone data signatures is first applied to CDR to achieve a coarse-grained classification of the city. Then, inside each retrieved cluster, the same technique is re-applied over signatures built from taxi pick-ups and drop-offs, as obtained from the GPS data. Such a combined use of mobile phone data and GPS traces has outperformed previous approaches when confronted with ground-truth information and allows characterising land use in greater detail.

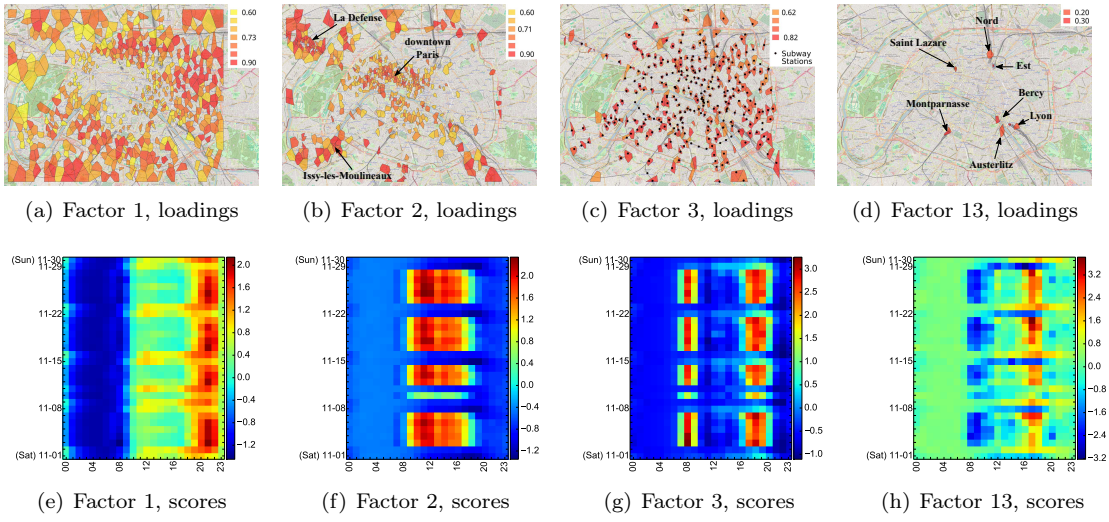
2.1.3 Spatio-temporal Profiling of Mobile Traffic Demand

Both the framework for mobile traffic temporal classification and the approach for urban fabric detection focus on one single dimension (either the temporal or the spatial one) to discover hidden structures in mobile phone data, by means of clustering. In the last part

of my postdoc, I focused on devising a novel solution for concurrently inspecting both the space and the time dimensions, which is a much more challenging problem. Also, clustering is traditionally limited by the fact that each element may belong to only one group, while there can be situations where an object has mixed behaviours and should be, therefore, partially associated with multiple classes. Such situations are not detectable with simple clustering-based techniques.

Therefore, I've explored the application to mobile phone data of a data reduction technique, known as Exploratory Factor Analysis (EFA). EFA is a well-established instrument in psychology research [119], but it was first applied to mobile phone data mining by our research group [10]. EFA aims at identifying, in a fully automated way, latent factors that cause the dynamics observed in the data. More precisely, given a set of random variables ($\mathcal{X}$) observed over a large population of samples, a reduced number of dimensions, or latent common factors ($\mathcal{F}$), is assumed for decomposing the vector of observed variables into the product of the hidden factors ($\mathcal{F}$) and a matrix of loadings (Λ). The loadings describe the importance of each observed variable with respect to the assumed factors.

When tailored to the specific use case of mobile traffic classification, EFA offers the possibility of exploring the space and time dimensions of the data at once, depending on how the observed variables are selected. This yields two significant advantages. First, the same methodology can be cast to recognise factors that are temporal or spatial in nature, solving the two problems of network activity temporal profiling (time intervals modelled as the EFA variables) and urban fabric detection (base stations modelled as the EFA variables). Second, our proposed EFA methodology allows immediate extrapolation of the structures hidden in the secondary dimension of both problems above. In other words, it provides, at no additional cost, knowledge of the spatial patterns that characterise each network activity profile, and of the precise temporal dynamics that distinguish each land use. This plays an important role in the interpretation of classification results.



(a) Factor 1, loadings (b) Factor 2, loadings (c) Factor 3, loadings (d) Factor 13, loadings

(e) Factor 1, scores (f) Factor 2, scores (g) Factor 3, scores (h) Factor 13, scores

Figure 2.5: Land use detection. (a)-(d) EFA of the total communication activity (sum of incoming/outgoing calls and SMS) in the Orange-2014 dataset. Loadings of the 1596 Voronoi cells (*i.e.,* EFA variables) on four (out of fourteen) representative classes (*i.e.,* EFA factors). (e)-(h) Thurstone's scores of the 91×24 hours (*i.e.,* EFA samples) on a selection of the 16 classes (*i.e.,* EFA factors). Figure best viewed in colours.

In Fig. 2.5, we report the results obtained via EFA (cast for solving the urban fabrics detection problem) when classifying the Orange data available for the city of Paris. First, the approach shows results that are comparable to our previous results on urban fabric classification in terms of spatial partitioning (compare Fig. 2.5(a):(e) with Fig. 2.3(f), Fig 2.4(b) and

Fig 2.4(d)). Second, the methodology automatically outputs the temporal classification (in Fig. 2.5(e):(h)), which provides insights for understanding how the factors are characterised with respect to time. Finally, the methodology allows the identification of areas exhibiting mixed usages, *i.e.,* with high loadings on multiple factors.

2.2 Origin-Destination Flows at Regional Scale via Network Signalling Data

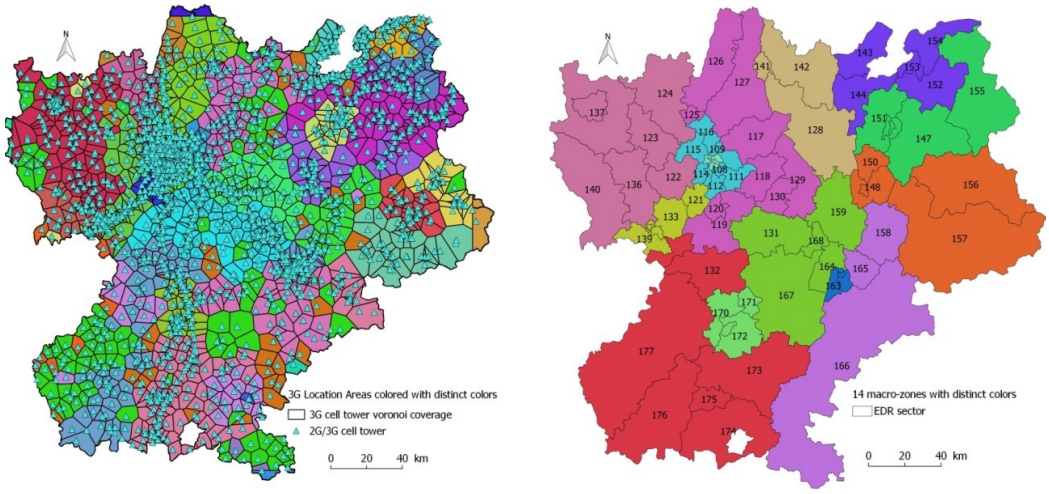
Spatio-temporal data are extremely valuable for studying human mobility for transportation and urban planning purposes [120, 121]. At large scales (*e.g.,* regional levels), traditional approaches rely on household travel surveys to collect mobility data, typically recording one day of travel diaries per household. However, surveys are increasingly confronted with issues during the sample construction phase [18], declining response rates [122], and unreported trips [123], all of which reduce the quality of the mobility indicators derived from them. Additionally, travel surveys typically involve high costs that limit their frequency (once or twice per decade) and prevent tracking the dynamics of population mobility over time. While travel surveys still provide highly useful data for constructing behavioural transport models (*e.g.,* route and transportation mode choice models) and coarse estimations of people's mobility in a city, they are less suitable for deriving reliable and time-varying mobility indicators due to the aforementioned issues. This is particularly true for OD matrices, whose accuracy and completeness can be severely compromised by the quality of the household surveys used.

Data collected from cell phones have become one of the most important new data sources to study travel behaviour [124]. Their attributes, such as large coverage of the geographic area, significant penetration in population and highly detailed location information, have attracted researchers to analyse them to support transportation studies. Several researches have been conducted to use different types of mobile phone data (*e.g.,* CDR, cellular network data); but, few have attempted to validate the results with external sources due to the different nature of mobile phone footprints. Yet, the validation process allows to identify possible biases and to gain a clearer idea of their potential. Moreover, the quality and accuracy of data is essential to ensure that investment or transport policy decisions are based on reliable analyses. Therefore, considerable efforts are needed to pre-process mobile phone data and to validate the related research outputs.

We pioneered the exploration of NSD for the accurate estimation of both static and time-varying OD matrices. We worked on a small (and limited) test dataset, related to one typical working day from 2017 (Thursday, June 1st 3:00 to Friday, June 2nd 3:00) provided by Orange Innovation. This work was performed in the context of Mariem Fekih's PhD thesis, in collaboration with the LAET laboratory of ENTPE and CNRS, the Hasselt University, Belgium, and the University of Bourgogne Franche-Comté, France. The primary objective of this study was to assess the capability of NSD to capture the major spatio-temporal mobility patterns of a mobile-device-equipped population at a large scale, by producing realistic static regional OD matrices. An additional objective was to validate the consistency of the reconstructed matrices with the ones obtained from household travel surveys, used as a reference. It was paramount to confirm this hypothesis before exploiting massive NSD for reconstructing mobility information at much finer spatio-temporal granularity based on the inference of individual GPS-like trajectories of human mobility (Chapter 3).

2.2.1 Static OD Matrix Estimation

To achieve the aforementioned objectives, we developed a simple methodological chain aimed to derive OD matrix via pre-processing of raw NSD data, and performed an extensive validation of the obtained results using the conventional OD matrix generated from a local travel



(a) Location of the cell towers and their Voronoi tessellation

(b) The 77 sectors of the EDR (labelled with unique numerical identifiers) and their aggregation in 14 macro-zones (represented with different colours)

Figure 2.6: Cell tower locations (a), and (b) EDR sectors in the region of interest (Rhône-Alpes).

survey related to the region of Rhône-Alpes, France. The explored NSD dataset consisted of 2G and 3G individual signalling records[5] of over two million anonymous mobile phone users in June 2017, totalling about 300 million records of device transactions. Concerning the spatial dimension, this dataset covered the whole extent of the Rhône-Alpes region in France, with 2,230 cell towers in the study area, and each cell tower possibly handling several antennas.

Fig. 2.6(a) presents the cellular network coverage within the Rhône-Alpes region and the aggregation in 3G Location Areas. The Rhône-Alpes household travel survey, called Enquête Déplacements Régionale (Regional Travel Survey) (EDR) 2015, contained data from 37,450 individuals aged over 11 years, identifying 143,000 trips. The sample was constructed using geographical stratified random sampling of the population of 77 zones in the region, graphically reported in Fig. 2.6(b). These 77 zones were spatially aggregated into 14 macro-zones represented in Fig. 2.6(b) with different colours. The survey contains socio-demographic characteristics and trip information, including transport mode, start and end time, activity at origin and destination, and location. Lyon is the largest metropolitan area in the region, and the survey only refers to working day trips during late autumn to early spring.

The Framework: Basic Definitions and Main Steps

The proposed framework, graphically shown in Fig. 2.7 and extensively described in [125], aggregates and transforms individual mobile phone signalling data (also called NSD user trajectory) into an OD flow matrix, formally denoted as $\mathcal{M} = \{\delta(z_i, z_j) \mid (z_i, z_j) \in Z \times Z\}$, with Z representing the set of 14 macro-zones, and $\delta(z_i, z_j)$ being the expanded daily count

[5]It is worth underlining that this first test release of NSD did not include any 4G communication data, which as from 2019, constitutes more than 50% of the worldwide mobile telecommunication technologies market. As detailed in Chapter 1, the absence of 4G data represents a substantial limitation in terms of richness of spatio-temporal information, thus further explaining the rather coarse grain of the results obtained in this preliminary study.

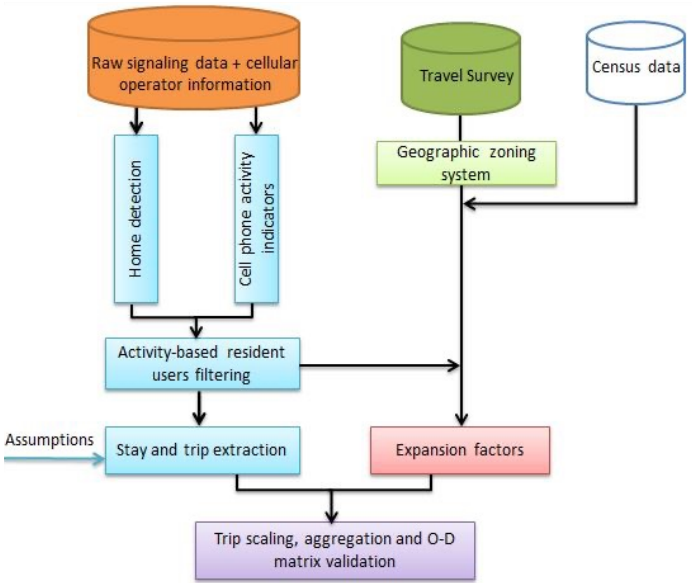


Figure 2.7: The OD matrix reconstruction framework.

of trips estimated via NSD between zone z_i and zone z_j[6]. The framework operates according to the following main steps:

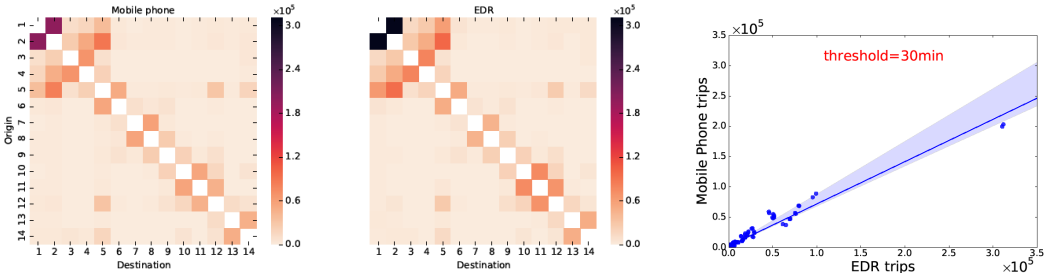
1. Filtering of (mobile) robot devices and identification of the most frequently observed cell tower to detect the *home location* of each user device[7] via time-based heuristics;
2. Identification of target resident users based on a condition of minimum required mobile phone activity (*e.g.,* each resident user should have a mobile event at least every 3 hours);
3. Estimation of expansion factors for each zone $z \in Z$ to scale the detected population of target resident users to the full population of travellers in the region of interest;
4. Trip detection and scaling up of the detected trips according to the estimated expansion factors and reconstruction of the final OD matrix $\mathcal{M}$.

It is worth remarking that the approach solely leverages NSD for user filtering and trip extraction, while it depends on travel survey and census data in relation to zoning (for spatial aggregation) and trip scaling (for determining the set of expansion factors), respectively.

The adopted definition of trips for NSD is based on the standard for travel survey introduced by CEREMA (the French national agency for transport network and urban planning) [126], which is at the core of the considered EDR survey. A trip is defined as any movement between pairs of *stationary activities*, each performed at a given zone. Thus, for a given user u, the whole set of individual NSD trajectories is scanned to identify consecutive observations on the mobile network of user u at the same zone for a minimum time threshold τ. For the choice of τ, we have considered multiple values (*e.g.,* 20, 30, 60 minutes) by

[6] We denote that in this work we did not take into account intra-zonal travel demand (*i.e.*, $z_i \neq z_j$ in the equation above). This restraining hypothesis was removed in the work described in Chapter 3.

[7] The framework assumes that a device corresponds to a single user. The two terms are used interchangeably in the following.



(a) The NSD reconstructed OD matrix. The matrix has been obtained with a threshold τ of 30 minutes for trip detection and aggregated at the level of the 14 macro-zones

(b) The EDR OD matrix (on 14 macro-zones)

(c) The regression line has the following equation with τ=30 minutes for trip extraction: $y_{ij} = 0.70 \cdot x_{ij} + 2193$, with y_{ij} representing the OD travel demand as seen via NSD data between zones z_i and z_j, and x_{ij} representing the EDR-based travel demand between the same pair of zones. The R^2 coefficient equals 0.96

Figure 2.8: Graphical comparison between the NSD-reconstructed OD matrix (a) and the EDR OD matrix (b). (c) Linear regression of the NSD OD matrix with respect to the EDR one.

taking into account the size of the zones (average area of EDR sector is 582 km^2). Based on such approach, each NSD user trajectory is ultimately converted into an ordered sequence of stationary activities $(A_1, A_2, ..., A_k)$, each associated to a given zone. The converted user trajectory can thus be parsed to extract $k-1$ trips, each defined as a tuple (u, O_i, D_i) where u denotes the given user, O_i the zone of the generic stationary activity A_i and D_i the zone associated to the following stationary activity A_{i+1}, being O_i and D_i the origin and the destination zones of the i^{th} trip of user u.

Expansion factors have been determined on a per zone basis, by considering population data from census data provided by the INSEE (Institut national de la statistique et des études économiques), aggregated at the level of the 77 sectors of the EDR survey. The expansion factor $\epsilon(z)$ for the generic zone z is calculated according to the formula: $\epsilon(z) = \frac{P(z)}{R(z)}$, where $P(z)$ and $R(z)$ represent the amount of population older than 11 in zone z from census data and the total amount of resident users detected via mobile phone data in the same zone, respectively. Expansion factors are then associated to users based on the detected home location, *i.e.,* $\epsilon(u) = \epsilon(z) \mid z = h(u)$, where $h(u)$ denotes the function that associates a home sector to user u. Finally, the static OD matrix $\mathcal{M}$ is computed per pair of origin and destination zones (z_i, z_j): the expanded trips between the given pair are summed up over the whole user basis, *i.e.,* $\delta(z_i, z_j) = \sum_{u \in U} \epsilon(u) \cdot n_u(z_i, z_j)$, with $n_u(z_i, z_j)$ representing the total number of trips detected for user u between zones z_i and z_j.

The main outcome of the framework, *i.e.,* the OD matrix $\mathcal{M}$, is visualised in Fig. 2.8(a) together with the OD matrix corresponding to the EDR survey, shown in Fig. 2.8(b).

The visual comparison of the distribution of OD trips from mobile phone and EDR data shows very similar shapes even though the total numbers of trips are different. The resemblance of the two matrices has been quantitatively confirmed by considering the Spearman's rank correlation and the R^2 coefficient of determination for a linear regression of the NSD-reconstructed matrix as a function of the EDR one, at macro-zone level (*i.e.,* 14 macro-areas). We obtained a satisfying result of $\rho = 0.95$ with $(p < 0.0001)$ and an R^2 coefficient of 0.96, which provides a high-confidence indication that the distributions of OD flows are similar. Hence, although both signalling data and survey-based matrices are collected using different techniques and technologies, and despite the limitation of the test dataset used for this study (*e.g.,* missing 4G data, one-day-only), they appear to resemble well, at least at the level of the 14 identified macro-zones, thus demonstrating the feasibility of a large-scale

Origin-Destination survey based on alternative passive real-world data, and motivating the research on the topic detailed in the next sections.

2.2.2 Dynamic OD Matrix Estimation

The preliminary study described in the previous section encouraged a deeper exploration into the use of NSD for OD matrices reconstruction by focusing on the challenge of using mobile phone data to generate time-dependent travel demand matrices.

Understanding the dynamics of human mobility patterns is a core notion in transportation studies related to, *e.g.,* traffic congestion management and transport infrastructure planning [127]. However, traditional household travel surveys typically record only one or a minimal number of days of travel diaries per household. Moreover, the high costs of conducting travel surveys restrict their frequency and prevent from following the mobility dynamics. In this respect, they do not allow considering fine-grained and continual temporal analysis of, *e.g.,* the hourly, weekly or special-events related variability of trips [128].

Communication data have instead the potential to describe mobility trends over long periods and with various time resolution scales, due to the ubiquitous and continuous use of mobile devices by the moving population. Hence, it appears reasonable to leverage mobile phone data to generate dynamic origin-destination flows by assigning the inferred trips into target time windows. Some studies have been conducted along this line to extract dynamic trip metrics using different forms of mobile phone data. Most of these studies have explored CDR and developed techniques to figure out the temporal distribution of user trips in limited geographical areas. However, it has been shown [129, 130] that these methods perform rather poorly, especially in urban zones, due to the very low spatio-temporal resolution of CDR data.

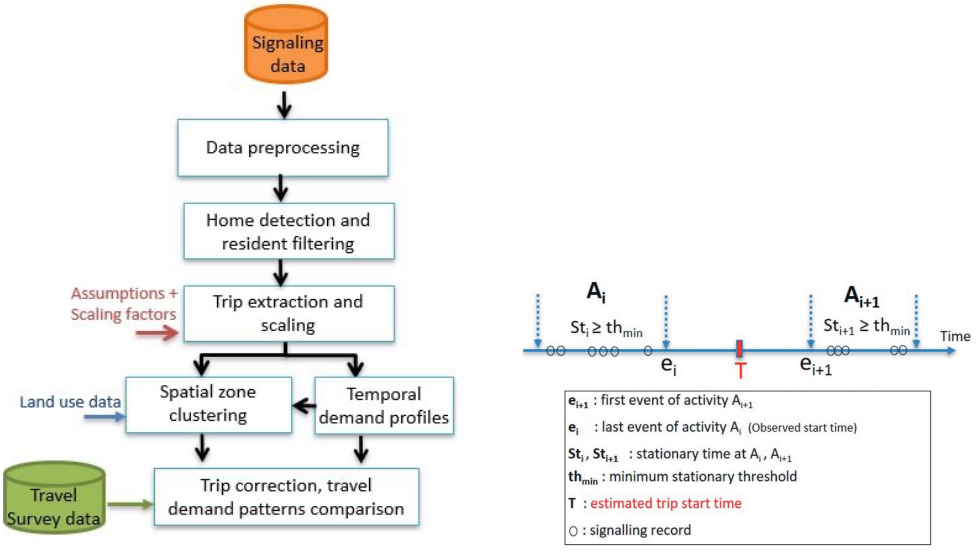
To answer the challenge of reconstructing dynamic OD flow at large-scale, we enhanced the workflow described in Sec. 2.2.1 for NSD. Specifically, we included a novel temporal component for estimating a start-time reference denoting the likely beginning of each observed trip.

The Framework: Basic Definitions and Main Steps

The extended framework is reported in Fig. 2.9(a), together with a schematic representation of the approach introduced to infer the trip start time (Fig. 2.9(b)). In the following, we focus on the description of the temporal component, while the full approach, largely based on the framework described in Sec. 2.2.1, is detailed in [131].

First, let us recall that in the approach of Sec. 2.2.1, each user's trip is reconstructed from NSD as an ordered sequence of stationary activities, *i.e.,* $(A_1, A_2, ...A_k)$. In this representation, no intermediate point (*i.e.,* way-point) between consecutive stationary activities is preserved from the original NSD user trace. This was a deliberate choice due to the reported limitation of the dataset (no 4G data). Moreover, this approach did not account for a proper characterisation of the positive bias existing between the last observed event of a stationary activity and the actual beginning of the trip. The problem of quantifying and removing the bias on trip start (and end) time will be discussed in detail (see Sec. 3.1) in the work [132].

In the framework, the trip taking place between two consecutive stationary activities $(A_i,\ A_{i+1})$ is supposed to start at any specific time instant between the timestamp t_i of the last observed mobile network event associated to activity A_i and the timestamp t_{i+1} of the first mobile network event associated to activity A_{i+1}. The absence of any way-point between the two activities does not allow, in fact, any further assumption on the moment the trip has started, apart from considering that the start time is uniformly distributed in the time window $[t_i,\ t_{i+1}]$. An enhanced approach relying on way-points will be described in Sec. 3.1. This procedure allows removing the biases on trip start and arrival times and



(a) The extended framework

(b) Start and end time trip imputation

Figure 2.9: Reconstruction of time-dependent OD matrices: (a) the extended framework; (b) The approach for imputing trip starting time.

makes possible a more robust estimation of the trip features. In the restrictive settings defined above, we were instead compelled to define the start time of a trip as the midpoint of the time interval separating the two activities, specifically $\frac{t_i+t_{i+1}}{2}$, which corresponds to the mean of the uniform distribution. Based on this approach, the trip can be finally represented as the tuple (u, t_i, O_i, D_i) where u denotes the given user, t_i the estimated start time of the trip, O_i the zone of the generic stationary activity A_i and D_i the zone associated to the following stationary activity A_{i+1}, being O_i and D_i the origin and the destination zones of the i^{th} trip of user u.

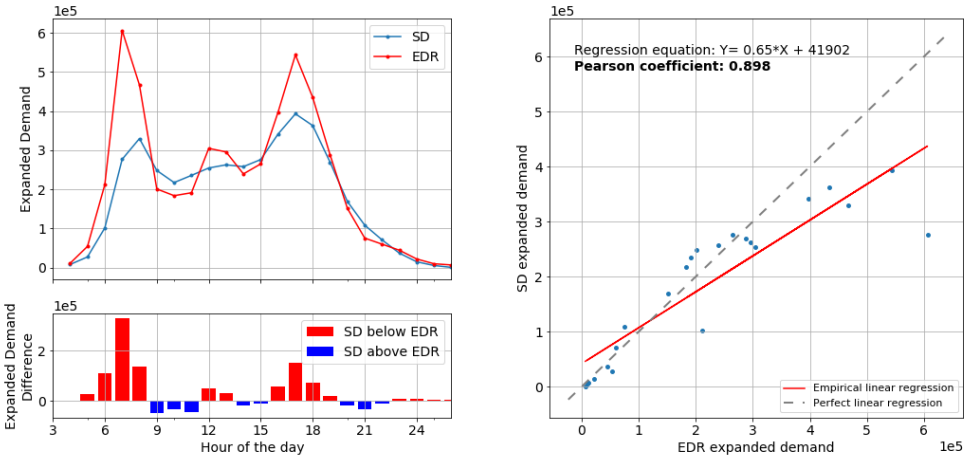
Thus, the final dynamic OD matrix can be defined as a temporally-sorted sequence of timestamped OD matrices $(\mathcal{M}(t) \mid t \in T)$, with T the ordered set of time slots considered within a typical day (*e.g.,* 24 hourly time slots) and the generic timestamped OD matrix $\mathcal{M}(t) = \{\delta(z_i, z_j, t) \mid (z_i, z_j) \in Z \times Z\}$. In the previous equation, $\delta(z_i, z_j, t)$ represents the number of expanded trips observed between zones z_i and z_j, *i.e.,* $\sum_{u \in U} \epsilon(u) \cdot n_u(z_i, z_j, t)$, with $n_u(z_i, z_j, t)$ the total number of trips detected for user u between zones z_i and z_j with start time within time slot t.

Empirical Evaluation

As a case study of the extended framework, we applied the methodology to the same NSD records of Sec. 2.2.1 related to the city of Lyon, France (*i.e.,* from 1st June 3:00 to June 2nd, 2017 3:00). As a comparative source of data, we used again the regional travel survey EDR 2015 conducted in the Rhône-Alpes region. The dynamic OD matrix was reconstructed at the spatial resolution of the 77 EDR sectors.

The analysis of the reconstructed dynamic OD matrix $\mathcal{M}(t)$ unveils interesting insights about the potential and the limitations of NSD for dynamic travel demand reconstruction.

The hourly total travel demand profile (*i.e.,* the number of hourly trips generated from all zones) is reported in Fig. 2.10(a) for both signalling and survey data. The total demand observed from signalling data (*i.e.,* SD) appears to be lower than the one reported in the



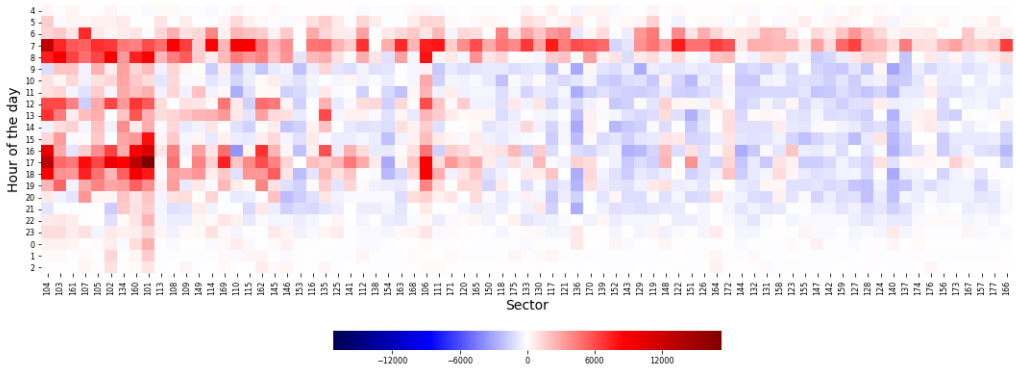
(a) (b)

Figure 2.10: (a) Temporal demand profile from signalling data (SD) and survey data (EDR) and demand difference between EDR and SD (red bars indicate an hourly EDR demand higher than the SD one, and blue bars otherwise) (b) correlation between the hourly demand estimations from SD and EDR.

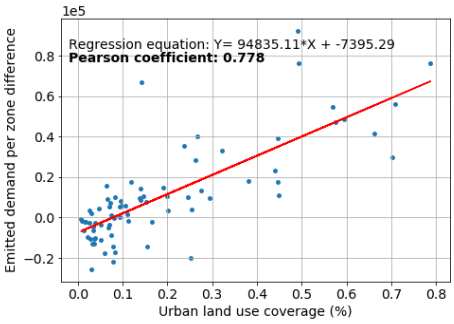
survey (*i.e.,* EDR), with less sharp morning and afternoon peaks, as clearly shown by the hourly difference ($\mathcal{M}_{EDR} - \mathcal{M}_{SD}$) reported in the bottom part of the same figure. This result could be explained by the existence of a large fraction of users in the mobile phone data, referred to as *static people* in the following, for whom it is possible to detect the home sector, but no trip can be observed, as the associated stationary activities produced are all (or mostly) located at the home sector. The proportion of such static people in our mobile phone dataset amounts to 46%. Travel demand reconstructed via NSD is unavoidably correlated to the number of events generated by resident users. Therefore, even though a certain portion of the identified static users could actually be stationary (*e.g.,* elderly people), it appears highly likely that another large portion of them is wrongly seen as static. The framework fails in associating trips to such sub-population, probably due to their very low mobile phone activity (*e.g.,* during morning hours). Such reduced device usage inevitably leads to an underestimation of the travel demand, which appears to be especially important during morning time, and requires a proper bias removal procedure. However, despite this underestimation, the hourly global demand profiles estimated from both data sources are highly correlated (Pearson coefficient equal to 0.898), as shown in Figure 2.10(b). This confirms that signalling data can provide a travel profile comparable to the well-known typical demand profile for a working day.

The analysis of the differences between dynamic OD matrices generated via signalling and survey data was performed not only from a temporal perspective but also from a geographical point of view. Fig. 2.11(a) shows the number of trips emitted from each sector of the study area, over the 24-hours reference period (*i.e.,* trips are counted with respect to the sector location of the trip's origin and the related start time). The heatmap provides a spatio-temporal representation of the absolute difference between the demand generated by each zone from the survey and mobile phone data. Sectors are sorted from left to right by decreasing density of urban land use, as retrieved from the *CORINE Land Cover 2012*[8] dataset. In highly-dense urban zones (on the left of Fig. 2.11(a)), the demand is higher with EDR survey compared to that reconstructed via mobile phone, regardless of the hour of the day. On the other hand, the hourly travel demand during the morning peak period is higher

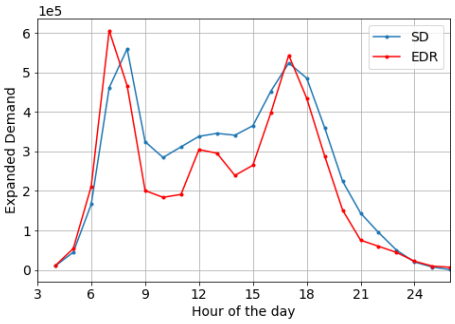
[8] https://www.statistiques.developpement-durable.gouv.fr/corine-land-cover-0



(a) Spatio-temporal distribution of the emitted demand difference between survey and signalling data



(b) Correlation between the emitted demand difference and urban land use percentage per sector



(c) Hourly demand distribution of EDR and Network Signalling Data (SD) after application of the de-biasing procedure

Figure 2.11: Analysis and correction of the biases in temporal NSD reconstructed travel demand.

in the survey compared to mobile phone regardless of the zone. This is confirmed by the plot reported in Fig. 2.11(b) showing that the emitted demand difference between survey and mobile phone is abnormally highly-correlated to the urban density of the sectors. In rural areas, we can reasonably assume that signalling records provide more consistent estimations since long-distance trips from/to these areas are typically better captured with mobile phone passive data than surveys and for a larger sample of the population [133]. Instead, in urban areas, the figures let us conclude that the proposed trip extraction method is unable to capture short-distance trips, which are expected to occur with higher frequency in urban areas than in rural ones. These preliminary analyses confirmed the existence of *systematic* spatio-temporal bias in the data.

Two simple heuristics were proposed to cope with these biases and used to correct the NSD-estimated travel demand. Concerning the geographical bias, we computed a spatial correction (multiplicative) factor, which was applied to the NSD-estimated travel demand on a per-zone basis. The correction factor was calculated using the following regression equation:

$$D_{Survey}(x) - D_{SD}(x) = 94835 \cdot U_{la}(x) - 7395 \tag{2.4}$$

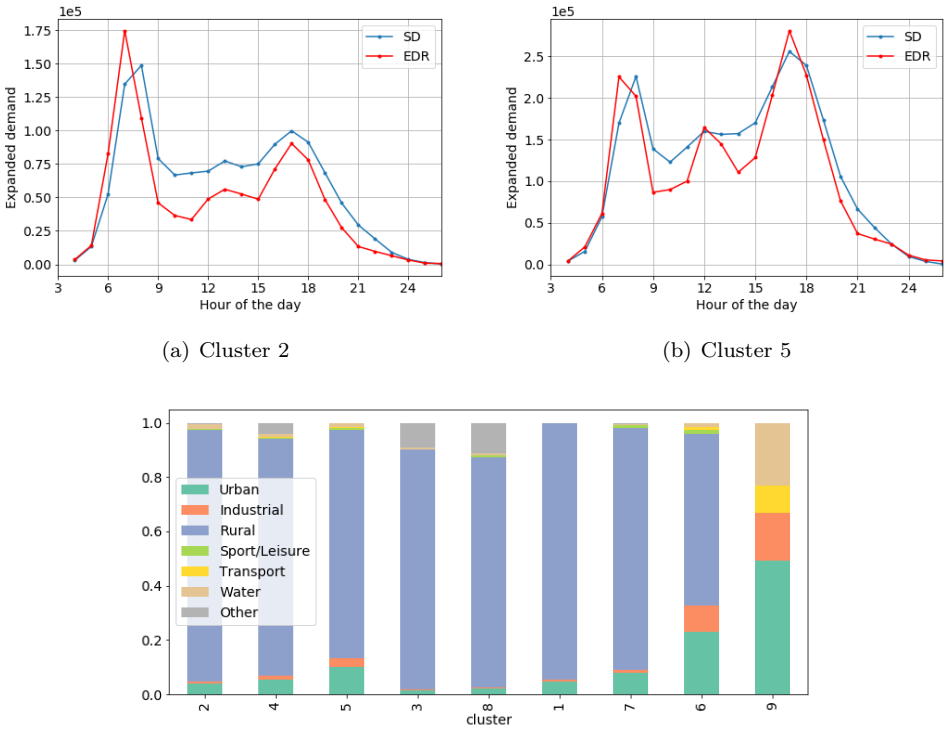
The regression equation was obtained by fitting a linear equation to the data visualised in Fig. 2.11(b), and expresses the demand difference between survey demand and NSD-estimated demand as a function of the urban land use density. The correction factor is therefore derived from Eq. 2.4 to remove the dependence on urban density from the differ-

ence. The factor is defined as follows:

$$S(x) = 1 + \frac{94835 \cdot U_{la}(x)}{D_{SD}(x)} \tag{2.5}$$

where $U_{la}(x)$ and $D_{SD}(x)$ represent, respectively, the urban density (as computed from land use data) and the emitted demand associated to each zone x (as estimated from the signalling data). By applying this correction factor, we can decorrelate the difference with respect to urban density.

Concerning the temporal bias, we have noticed that the observed underestimation (*i.e.,* , between 5 and 9 am) of the travel demand also affects the temporal distribution of Location Area Update (LAU) events (*i.e.,* a kind of communication passively generated by the device when changing the Location Area zone), which should not be normally affected by the time of the day. Therefore, to address this bias, a uniform correction factor has been applied on all mobile phone-based trips with a start time estimated during and around the morning peak period [5–9 am]. This factor has been calculated as the ratio of the afternoon and morning peaks in the LAU profile. Based on these considerations, the applied temporal correction factor results equal to a value of 1.3.



(a) Cluster 2

(b) Cluster 5

(c) The land use composition of clusters

Figure 2.12: Average temporal demand profiles for a selection of clusters (a-b). Land use distribution in all clusters (e).

Fig. 2.11(c) shows the final bias-corrected temporal profile comparatively to the EDR one, which clearly exhibits a higher correlation (Pearson coefficient equals 0.94) with respect to the survey-based one.

As a final validation step, we have developed a clustering methodology to identify groups of zones showing similar temporal patterns in terms of the emitted, NSD-estimated hourly demand[9]. The methodology detects a variety of temporal demand patterns within the study

[9]The clustering approach was in all similar to the one used for grouping MWS of mobile phone traffic

area, partially reported in Fig. 2.12. The analysis of the average NSD-estimated emitted demand in each cluster has been performed comparatively to the EDR one (Fig. 2.12(a)-2.12(b)), and by taking into account the land use composition in each of the identified clusters (Fig. 2.12(c)). The results of the clustering clearly show an agreement in terms of demand profile between the two sources, as well as easily explainable dynamics with respect to the land use configuration of each cluster. As an example, cluster 2 (Fig. 2.12(a)) is mostly associated with rural areas, featuring a strong residential land use with a limited presence of businesses or leisure activities. The emitted demand confirms this conclusion with a profile unequivocally describing a large number of people leaving this cluster to reach working places at the morning peak and coming back at the late-afternoon peak. Conversely, cluster 5 (Fig. 2.12(b)) covers a large amount of industrialised urban areas. These zones are characterised by a mixture of residential (high population density) and industrial/office fabrics: they tend to generate a high number of commuting trips in the morning but also attract a significant amount of demand from the surrounding areas. The latter will massively leave the cluster (thus generating trips) later on, as confirmed by the higher afternoon/evening peak.

These results confirmed the interest in using NSD for dynamic travel demand reconstruction but also highlighted the absolute need for more structured processes that might account for bias modelling, trajectory cleansing and reconstruction of missing observations, towards estimating finer-grained mobility indicators.

activity detailed in Sec. 2.1.2), with the only difference that it was applied to 24-hours time series of emitted travel demand.

Chapter 3

Zonal Traffic Speed and Urban Trajectory Reconstruction

The outcomes of the research described in the previous chapter demonstrated the possibility to leverage individual massive mobile phone data to estimate OD matrices at coarse regional scale, with accuracy consistent with that of traditional household travel surveys. Moreover, these activities provided evidence that passively collected mobile phone data could enable fine-grained inspection and continuous monitoring of human mobility.

This chapter describes the research conducted on mobile phone data to further advance the understanding established in Chapter 2. Specifically, this chapter focuses on achieving a higher level of spatial granularity in the estimation of traffic variables and mobility trajectories by referring to small areas of the urban territory, ranging in size from a few tens to hundreds of square meters. It also aims to capture these variables and mobility trajectories at a shorter temporal scale, in the order of minutes or tenths of minutes. Moreover, all the approaches developed in this chapter enable large-scale reconstruction of mobility information, *i.e.,* encompassing the whole urban area of a city, and, in some cases, even a whole region or country.

In this direction, as part of Manon Seppecher's PhD thesis, we investigated the possibility of using individual massive mobile phone data to estimate time-dependent zonal mean spatial speeds[1] of a partitioned urban road network. It is worth mentioning that, in this research, we did not explicitly work on NSD or CDR. Instead, we opted for an abstract data representation that we denoted as UAPD[2]. This representation can be customised to suit a variety of datasets, such as CDR, NSD, LBSN data, or any similar dataset generated by users who, while static or in motion, irregularly interact with a device to perform a task (such as communicating, posting content on social media, *etc.*). The main outcomes of this research are reported in Sec. 3.1.

In parallel to this research activity, we had the opportunity to work on an extensive, complete, and exceptionally large-scale NSD dataset, already presented in Sec. 1.1.2 to showcase the increased accuracy, from both the temporal and spatial points of view, of NSD in comparison to CDR. This dataset was collected and made available by Orange Innovation and provided a realistic representation of what a network operator could collect with state-of-the-art network event probes[3]. The unique features of such datasets enabled

[1]The zonal mean spatial speed represents an essential variable for macroscopic traffic modelling, according to the theory of the Macroscopic Fundamental Diagram [134].

[2]This choice was also imposed by a delay in the acquisition of real CDR data, which were supposed to be available in the context of Seppecher's PhD thesis, and the impossibility of using Orange data which was not a partner of the project supporting these activities. The experimentation was therefore conducted on a UAPD dataset obtained from a massive GPS data, artificially biased to resemble CDR and NSD data.

[3]For more details on the probe technology and approach adopted by Orange Innovation on the French national network to collect NSD datasets, we refer the interested reader to [135].

us to investigate the problem of utilising NSD to reconstruct human mobility trajectories in urban environments with spatio-temporal accuracy of a few hundred meters, *i.e.,* comparable to the one that can be obtained with GPS data. Solving this problem at the scale of a whole city, and for a large sample of the city population[4], allows estimating high-quality and highly-detailed mobility patterns, leading to a better understanding of human mobility in urban settings. In Sec. 3.2, we describe the framework implemented to accurately reconstruct urban mobility trajectories, while in Sec. 3.3, we present the enhanced solution adopted to perform map-matching of the reconstructed trajectories on a multi-modal transport network.

The increased accuracy and resolution of the retrieved mobility information, although beneficial from the perspective of city management and planning, also raises fundamental questions about users' privacy. It could become in fact relatively easy to retrieve sensitive information about people, such as the precise location at a specific time, their home and work positions, and even to unveil a user's identity when such sensitive information are cross-referenced with other datasets. Therefore, we also dedicated significant research efforts to the problem of devising anonymisation solutions aimed at reducing the sensitivity of the reconstructed mobility indicators, while preserving the spatial and temporal granularity of the mobility patterns at the finest possible level of detail. Sec. 3.4 reports on the methodology developed to anonymise OD matrices derived from massive spatio-temporal trajectories of human mobility.

The activities on mean speed dynamics estimation have been conducted in collaboration with CITEPA, in the context of Manon Seppecher's PhD thesis and the Green City Big Data project (Jan. 2018 - Apr. 2022). The analysis and processing of massive NSD datasets to enable the estimation of fine-grained mobility patterns has been one of the central topics of the ANR JCJC PROMENADE (Feb. 2019 - July 2023) project, which I have led. The framework presented in Sec. 3.2 has been a major contribution of Loïc Bonnetain's PhD thesis, which I co-directed, and the result of a successful collaboration with INRIA CITI Lab, CNR Italy, and Orange Innovation. The framework has also been recently patented [136][5]. The research activities related to the anonymisation of mobile phone trajectories and of the derived mobility indicators have their origins in the work described in [137], performed as part of my post-doctoral activities. In such work, we introduced a novel algorithm, called GLOVE, for the anonymisation of highly detailed trajectories derived from mobile phone data. Based on these results, a novel approach to the anonymisation of mobile phone signalling traces has been proposed in the context of the ANR PRCE MOBITIC (Jan. 2020 - Dec. 2024) project, and are part of Benoit Matet's PhD thesis, which I currently co-advise in collaboration with researchers from the GRETTIA laboratory of University Gustave Eiffel. They have also been part of a collaboration with the IMDEA Networks Institute of Madrid.

This chapter includes content from the following papers:

1. M. Seppecher, L. Leclercq, A. Furno, D. Lejri, and T. V. da Rocha, "Estimation of urban zonal speed dynamics from user-activity-dependent positioning data and regional paths", Transportation Research Part C: Emerging Technologies, vol. 129, p. 103183, 2021.

2. L. Bonnetain, A. Furno, N.-E. El Faouzi, M. Fiore, R. Stanica, Z. Smoreda, and C. Ziemlicki, "Transit: Fine-grained human mobility trajectory inference at scale with mobile network signaling data", Transportation Research Part C: Emerging Technologies, vol. 130, p. 103257, 2021.

3. L. Bonnetain, A. Furno, and N.-E. El Faouzi, "Multi-modal fine-grained map-matching of mobile phone network signaling data in urban areas", 101st Transportation Research Board Annual Meeting (TRB), 2022.

[4] Around 35% to 40% of people in France are customers of Orange. This represents a significantly larger sample compared with 1% of persons reached by mobility surveys.

[5] https://data.inpi.fr/brevets/FR3125197

4. B. Matet, A. Furno, M. Fiore, E. Côme, and L. Oukhellou, "Adaptative generalisation over a value hierarchy for the k-anonymisation of origin–destination matrices", Transportation Research Part C: Emerging Technologies, vol. 154, p. 104236, 2023.

These results are currently supporting further ongoing research related to resilience-driven design and reconfiguration of multi-modal transport networks as well as multi-source data fusion for travel demand anomaly detection and synthetic population generation, which will be described in Chapters 4 and 6, respectively.

3.1 Estimation of Mean Spatial Speeds in Urban Areas

Estimating the dynamics of traffic speed in urban road networks is crucial for many applications, including traffic simulation and control, or greenhouse gas emission estimation [138]. Targeting speeds from irregular and low-frequency positioning data is a challenging task. The traditional bottom-up speed estimation methods from GPS floating vehicle tracks [139, 140] rely on averaging individual speeds calculated at the road segment level. These approaches cannot be easily transposed to mobile phone data such as CDR or NSD. As highlighted in Chapter 2, the accurate estimation of start and end times of individual trips – and thus travel times and speed values – from mobile phone trajectory data is problematic even at a coarse commune scale due to the inherent spatio-temporal biases of mobile phone data. However, these data provide an unprecedented spatio-temporal coverage of the urban territory that cannot be efficiently achieved with other sources, such as GPS or pollution sensors, thus justifying significant research efforts in the investigation of methods for reliable speed estimation with UAPD.

One of the primary objectives behind the Green City Big Data project was in fact to develop methods for estimating traffic variables and assessing, with the highest possible level of details, traffic-related pollutant emissions at the scale of a whole metropolitan city. To that end, we decided to approach this issue from an intermediate regional scale suitable for precisely studying the dynamics of large urban road networks, but coarser than the road segment level to properly handle the sparsity of mobile phone data. Based on such intermediate scale, we identified specific emission models (*i.e.,* the COPERT [141]) tailored to the data characteristics, the urban scale of emission assessment, and the decision-making context in which the tool was intended to operate. The COPERT model leverages mean traffic speeds and vehicle features, such as motorisation and emission standards, to quantify the emissions produced per unit of distance travelled. Consequently, estimating global emissions at a regional and urban scale requires multiplying these factors by the total traffic volume[6]. Specifically, the emissions of pollutant k can be estimated using:

$$E_k = TTD \cdot F_k(V) \tag{3.1}$$

where TTD represents the Total Travel Distance (in km), V stands for the traffic speed (in km/h), and $F_k(V)$ denotes the emission factor (in g/km).

Within the context of Manon Seppecher's thesis, we developed a comprehensive modelling chain that uses UAPD data, road network data, and census and survey data as inputs. This chain produces the traffic variables required for emission estimation, including traffic speeds, travelled distances and traffic volumes. The chain is structured around the speed estimation process and employs different approaches for reconstructing the traffic volumes related to two distinct user categories, *i.e.,* regular and non-regular travellers of the selected urban territory. In the following, we only focus on the framework for the estimation of the mean spatial speeds. The methodologies related to the estimation of traffic volumes for the different user categories and the overall framework for pollutant estimation can be found by the interested reader in [142, 143].

[6] The total travelled distance TTD_r within a region r is linked to the traffic flow q_r in the region and the average trip lengths L_r in r according to the following equation $TTD_r = q_r \cdot L_r$ [134].

To estimate speeds, we considered $\mathcal{R}$ as the given regional partitioning of the urban road network, with r being the generic region within this partitioning. These regions must be characterised by homogeneous city fabric, demography, road network topology, and, most importantly, traffic dynamics. Homogeneity of traffic dynamics is an essential requirement for a robust estimation of aggregate traffic variables, as reported by the literature on the Macroscopic Fundamental Diagram (MFD) theory [144], which underlies large-scale traffic modelling and simulation. Following the network partitioning guidelines provided by this literature, one can typically divide a city into a set of regions ranging from 5 to 20. This new spatial scale determines the final spatial resolution of traffic speed estimates. Therefore, it must be adapted to the precision requirements of the case study and to the resolution of the available data, as mentioned above. Concerning the temporal scale, we assumed time to be discretized into equal-length time slots t, drawn from a temporal set $\mathcal{T}$, which corresponds to the entire observation period (*e.g.,* one day). The length of the time slots imposes the time granularity of the speed estimates and must be chosen accordingly. In must be small enough to reproduce the rapidly changing speed dynamics during peak hours, but sufficiently large to ensure an adequate size of the trip sample population. We chose 15-minutes time slots in this study, as commonly used in the literature.

Under these hypotheses, our aim was thus to accurately estimate the mean spatial speed V_r^t for each $r \in \mathcal{R}$ and $t \in \mathcal{T}$, which is defined as follows:

$$V_r^t = \frac{TTD_r^t}{TTT_r^t} = \frac{\sum_i L_r^{i,t}}{\sum_i T_r^{i,t}} \tag{3.2}$$

In this equation, TTD_r^t represents the sum of the individual distances (*i.e.,* network lengths) travelled in region r during time slot t, *i.e.,* $L_{r,t}^i$. Meanwhile, TTT_r^t represents the sum of the individual travel times in region r during time slot t, *i.e.,* $T_r^{i,t}$.

The fundamental principle behind the speed estimation method we proposed is that the overall sample size of the data can compensate for the low data quality at the individual trip level. The method relies on the fusion of individual trip information and statistical considerations to provide a reliable traffic speed estimation.

Basic Definitions

Similarly to the approach followed in Sec. 2.2, we adopted a definition of *trip* based on the observed individual UAPD trajectory. Specifically, the trip is defined as the mobility phase of a user between two consecutive stationary activities, named *stays* in the following. We remind here that an important distinction must be made between the observed trip departure and arrival times and the exact (but unknown) ones, as the varying communication rates of users provide sparse information on their mobility. Based on these considerations, we used the following definitions:

- The *regional path* is defined as the sequence of the successive regions travelled by a user in a trip from its origin to its destination [145, 146]). It corresponds to a coarser representation of the path followed by the traveller with respect to the road segment scale, based on the considered partitioning into regions of the urban road network.
- We call *trip* the ternary structure defined by a regional path, an observed travel time and an observed arrival time. Differently from the approach described in Sec. 2.2.2, this definition of trip does not ignore way-points, *i.e.,* events observed for a given user between two successive stays. Intermediate user positions, represented as detection of the user at specific regions from the chosen partitioning, can provide precious information to infer accurate speed estimations.
- The *observed departure time* of trip i is defined as the time when the last static event

of the origin stay is observed. By definition, the observed departure time precedes the actual one. Thus, ϵ_d^i represents the positive bias between these two values.

- Reciprocally, the *observed arrival time* of trip i is defined as the time when the first static event of the destination stay is observed. By definition, the observed arrival time follows the actual one. Thus, ϵ_a^i represents the positive bias between these two values.
- The *observed travel time* T_{obs}^i of trip i is defined as the time elapsed between two consecutive stays, *i.e.*, between the observed departure and the observed arrival times. It is an overestimate of the actual travel time T^i.

Based on these definitions, we thus have:

$$\epsilon^i = T_{obs}^i - T^i = \epsilon_d^i + \epsilon_a^i \tag{3.3}$$

The Framework: Main Steps

In the following, we provide a summary of the speed estimatino framework (see Fig. 3.1) by focusing on the procedures for travel time bias removal and for speed estimation. The interested reader can refer to [132] for the complete description of the approach.

The framework assumes the availability as inputs of: *(i)* a network partitioning into a set $\mathcal{R}$ of homogeneous regions; *(ii)* a set of relevant regional trip length information. The estimation of trip length information must be performed beforehand from external data sources, on the same network partitioning that defines the adopted regional scale and paths. We relied on previous work for the definition of the network partitioning [147] as well as for the estimation of the regional trip lengths [146]. Specifically, this input corresponds to the average regional trip length in each region along each possible regional path. We assume this information to be available as a distance matrix $\hat{\boldsymbol{L}}$, whose rows (i) are the different possible regional paths, and columns (j) are the different regions resulting from the spatial tessellation. The cell value at (i, j) corresponds to the average distance travelled in the j^{th} region, when travelling along the i^{th} regional path P. It is equal to zero if the path P does not cross the j^{th} region. This matrix was assumed to be constant over time in our framework, but a time-dependent generalisation could be considered as well if such information can be derived independently from another dataset [148].

Travel Time Bias Modelling and Removal

We modelled the bias as an additive component on the exact travel time of any trip observed via UAPD[7]. Let P be a regional path and i be an individual trip observed along P. We thus have:

$$T_{P,obs}^i = T_P^i + \epsilon^i \tag{3.4}$$

where $T_{P,obs}^i$, T_P^i and ϵ^i are, respectively, the observed travel time of trip i along P, its exact travel time, and the corresponding travel time bias. The latter can be expressed as:

$$\epsilon^i = \epsilon_d^i + \epsilon_a^i \tag{3.5}$$

Estimating this bias on an individual trip basis is difficult. However, the estimation of its average is less challenging and can allow de-skewing, on average, the observed travel times. To compute the average bias, we proposed merging overlapping trips and averaging their observed travel times to build a unique aggregated biased travel time information per path and per time slot.

[7] Non-additive and more complex forms of bias could have been considered, but we opted for the most natural and simple representation in this work.

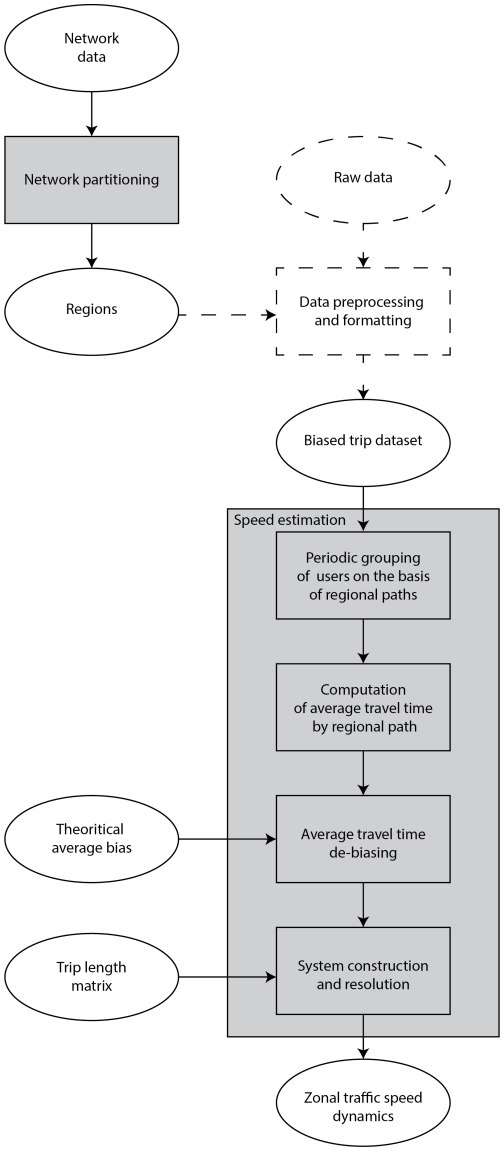


Figure 3.1: Methodological framework for regional speed reconstruction.

Let t represent a generic time slot, and let I_P^t be the set of overlapping trips along P that reach destination during t, with $n_{t,P} = |I_P^t|$. Averaging Eq. 3.4 over I_P^t gives:

$$\bar{T}_P^t = \bar{T}_{P,obs}^t - \bar{\epsilon}_P^t \tag{3.6}$$

where $\bar{T}_P^t$, $\bar{T}_{P,obs}^t$ and $\bar{\epsilon}_P^t$ are, respectively, the average actual travel time, the average observed travel time, and the average bias of trips from set I_P^t.

Eq. 3.6 can be simplified by assuming that the bias is independent of the trip path and time[8]. Thus, the distribution of individual biases ϵ^i can be modelled via a unique random variable X. The construction of such a model, and the estimation of its first moment

[8]The frequency at which mobile phone users are observed through the mobile network, *i.e.,* the inter-event time, is known to depend on the hour of the day [149]. In particular, longer inter-event times are observed during nighttime and early morning. This could make the bias, in turn, dependent on time. However, these results normally account for all individuals, including the ones that are static and sleeping, while, in this study, we were exclusively interested in the moving ones, for which it might be reasonable to make the simplistic hypothesis of time independence. The spatial independence of the bias is similarly debatable and

$\mu_X \equiv E(X)$, can offer an approximation of $\bar{\epsilon}^t_P$ allowing the de-skewing of $\bar{T}^t_{P,obs}$, provided that the sample of individuals associated with this time slot and path is large enough:

$$\bar{T}^t_P \approx \bar{T}^t_{P,obs} - \mu_X \tag{3.7}$$

Hence, if the departure and arrival temporal offsets ϵ^i_d and ϵ^i_a are themselves modelled by the same random variable Y, Eq. 3.5 gives $X = 2Y$. Now, as discussed in Sec. 2.2.2 but with a finer modelling taking into account way-points, the individual's departure time from a stay position can occur with a uniform probability between the pre-departure communication event and the post-departure communication event. The delay between these two events follows the distribution of the user's inter-event times, which can be assimilated to the population's inter-event times distribution for the sake of simplicity.

Mathematically, this means that departure bias follows a uniform distribution law bounded by the population's inter-event time distribution. Symmetrically, the same reasoning applies to the arrival bias. This modelling, focusing on the features of the whole population in terms of communication patterns instead of considering specific properties of the individual trip, represents a significant improvement with respect to the simple approach adopted in Sec. 2.2.2 to estimate the duration of trips from individual passively collected traces.

We modelled the population's inter-event time by considering an exponential law Z of parameter λ. While in the literature the inter-event time distribution is often modelled as a truncated power law distribution, we selected an exponential distribution for a matter of simplicity. It requires a single parameter λ directly linked to the distribution average. Enhancing this modelling choice represents a possible future research direction.

The considerations above leads to:

$$Z \sim Exp(\lambda) \tag{3.8}$$

$$Y|Z \sim U(0, z) \tag{3.9}$$

Hence, the probability density function of Z, and the conditional probability density function of Y given the occurrence of the value z of Z can be written as:

$$f_Z(z) = \lambda e^{-\lambda z} \tag{3.10}$$

$$\text{and } f_{Y|Z}(y \mid z) = \begin{cases} \frac{1}{z} & 0 \leq y \leq z, \\ 0 & \text{otherwise,} \end{cases} \tag{3.11}$$

It can be proved (see [132]) that the probability density function of Y is:

$$f_Y(y) = \lambda \int_0^{+\infty} \frac{e^{-\lambda(y+z)}}{y+z} dz \tag{3.12}$$

and that the first two moments of Y are:

$$E(Y) = \frac{1}{2\lambda} \qquad \text{and} \qquad V(Y) = \frac{5}{12}\frac{1}{\lambda^2} \tag{3.13}$$

Those results characterise the random variable Y which models the departure and the arrival offsets. This gives for $X = 2Y$:

$$\mu_X \equiv E(X) = 2E(Y) = \frac{1}{\lambda} \qquad \text{and} \qquad V(X) = 4V(Y) = \frac{5}{3}\frac{1}{\lambda^2} \tag{3.14}$$

It is important to highlight here the importance of the size of I^t_P to ensure that μ_X is representative of the average bias related to regional path P at time t: the larger the sample,

would deserve further study, since mobile phone or social network usages are known to be correlated with socio-demographic characteristics.

the more the theoretical average bias μ_X is representative of the sample's average bias. This is all the more important as with our bias model, as the variance of X increases with mean inter-event time $E(Z) = \frac{1}{\lambda}$:

$$V(X) = \frac{5}{3}E(Z)^2 \tag{3.15}$$

Eq. 3.15 shows that the larger the mean inter-event time is, the more scattered the trip bias distribution will be, and the more data per time slot and per regional path will be needed to ensure a reliable de-biasing process.

Regional Speed Estimation

The estimation of regional speed dynamics relies on the computation of mean regional path travel times. Specifically, the individual regional travel time from Eq. 3.4 can be expressed as the fraction of the distance travelled on trip i in r (i.e., $L^i_{P,r}$) over the mean spatial speed of i in region r (i.e., V^i_r), as described in Eq. 3.17.

$$\forall i \in I^t_P, \qquad T^i_P = \sum_{r \in P} T^i_{P,r} \tag{3.16}$$

$$T^i_P = \sum_{r \in P} \frac{L^i_{P,r}}{V^i_r} \tag{3.17}$$

Due to the temporal sparsity of the individual UAPD trajectories, the terms $L^i_{P,r}$, and V^i_r are considered unknown.

Although vehicles may experience different local and instantaneous speeds over an area, their average speeds depend mostly on overall traffic conditions, and mainly on the accumulation (*i.e.*, number of vehicles in the region). These average speeds show little scatter among individuals, and can be approximated by the mean spatial speed of all individuals travelling in the region. This observation has sustained the development of the MFD theory [144, 150] and motivates the choice of a network partitioning into sub-regions of consistent traffic dynamics. On this basis, it is possible to consider this equivalence for each trip i, being t the time slot trip i refers to, *i.e.,*

$$V^i_r = V^t_r, \quad \forall i \tag{3.18}$$

where V^t_r is the regional spatial mean speed at time slot t.

In Eq. 3.17, after summing on the $I_{t,P}$ trips, this gives:

$$\sum_{i=1}^{n_{t,P}} T^i_P = \sum_{i=1}^{n_{t,P}} \sum_{r \in P} \frac{L^i_{P,r}}{V^t_r} = \sum_{r \in P} \sum_{i=1}^{n_{t,P}} \frac{L^i_{P,r}}{V^t_r} \tag{3.19}$$

Eq. 3.19 can easily be rewritten as follows:

$$\sum_{i=1}^{n_{t,P}} T^i_P = \sum_{r \in P} \frac{1}{V^t_r} \sum_{i=1}^{n_{t,P}} L^i_{P,r} \tag{3.20}$$

$$n_{t,P} \bar{T}^t_P = \sum_{r \in P} n_{t,P} \frac{\bar{L}^t_{P,r}}{V^t_r} \tag{3.21}$$

$$\bar{T}^t_P = \sum_{r \in P} \frac{\bar{L}^t_{P,r}}{V^t_r} \tag{3.22}$$

Again, a significant advantage of this averaging process over the sample I_P^t is that the characterisation of individual regional trip lengths $L_{P,r}^i$ for any individual i becomes unnecessary. Instead, the sample mean value $\bar{L}_{P,r}^t$ turns out to be sufficient. On the condition that the sampling size is large enough, this can be replaced by its static estimate $\hat{L}_{P,r}$, drawn from the exogenous trip length matrix $\hat{\boldsymbol{L}}$ described above:

$$\bar{T}_P^t \approx \sum_{r \in P} \frac{\hat{L}_{P,r}}{V_r^t} \tag{3.23}$$

At this stage, the computed distance matrix is used to express, through Eq. 3.23, a relationship between the average travel time along regional path P at t, and the underlying, unknown mean traffic spatial speeds of the regions along P.

Although the actual average trip duration $\bar{T}_P^t$ is unknown, the knowledge of the average time bias μ_X allows to estimate it from the average trip duration, *i.e.*, $\bar{T}_{P,obs}^t$, observed from the UAPD data at each time slot t and for each path P. Based on Eq. 3.7, we thus get:

$$\bar{T}_{P,obs}^t - \mu_X \approx \sum_{r \in P} \frac{\hat{L}_{P,r}}{V_r^t} \tag{3.24}$$

Conversely, the constant distance parameters $\hat{L}_{P,r}$ can be drawn from the aforementioned estimated trip length matrix $\hat{\boldsymbol{L}}$. μ_X is assumed known as well. V_r^t are the only unknowns of the system. When applying in Eq. 3.24 the change of variable $x_r^t = 1/V_r^t$, we finally get the unbiased system:

$$\forall t, \quad S^t = \{\bar{T}_{P,obs}^t - \mu_X = \sum_{r \in P} \hat{L}_{P,r} x_r^t, \quad \forall P\}. \tag{3.25}$$

In Eq. 3.25, we name S^t the linear system composed of $|R|$ unknowns $(x_r^t, r \in R)$ and as many equations as the number of regional paths observed during the reference time slot t. The UAPD data analysis and the parameters extracted from the trip length matrix allow to fully characterise the system, which can be rewritten in matrix notation as follows:

$$\forall t, \quad S^t = \{\boldsymbol{T_{obs}^t} - \boldsymbol{\mu_X} = \hat{\boldsymbol{L}}^{|t} \boldsymbol{x^t}\} \tag{3.26}$$

where $\boldsymbol{T_{obs}^t}$ is the average observed travel time vector and $\hat{\boldsymbol{L}}^{|t}$ is the sub-matrix of $\hat{\boldsymbol{L}}$ restricted to the regional paths observed at time slot t.

Given that the number of regional paths will generally exceed the number of regions of the adopted partitioning, S^t is very likely over-determined. Consequently, the system will probably have no exact solution, but an approximated one can be calculated using regression analysis. To this purpose, we applied a non-negative least squares regression method to solve the system. In our case, the non-negative constraint allows for taking into account the non-negative nature of the zonal traffic speed.

Taking the reciprocal values of the solution vector $\boldsymbol{x_0^t}$ gives the optimal speed vector $\boldsymbol{v_0^t}$. This resolution process can be iterated throughout the whole studied time span to estimate the complete temporal speed trends.

Empirical Evaluation

We performed an extensive evaluation of the proposed framework by leveraging a cleaned and map-matched dataset of GPS trajectories that we down-sampled and artificially biased to mimic real-world mobile phone data. The GPS data were provided by a leader European navigation system provider[9] from October 2017 and September 2018 over the Greater Lyon

[9] https://be-mobile.com

(a) Map of the urban regions

(b) Map of the ring road regions. The ring road is divided into three zones, which are themselves separated into two according to the direction of traffic

Figure 3.2: Maps of the regions partitioning the city of Lyon, France.

area. The study area focuses on the urban area of Lyon and the neighbouring municipality of Villeurbanne, located inside Lyon's ring road. We partitioned this territory into 16 distinct regions, as displayed in Figure 3.2. The traffic variables were verified to be relatively homogeneous in each region [147]. The trips used in this study were extracted from five typical weekdays, *i.e.,* from Monday, February 12, to Friday, February 16, 2018. Being GPS data rather limited in sample size, we artificially extended the size of the dataset by duplicating each trip 100 times. Moreover, as few trips were observed at night-time in our dataset, the time span selected for our evaluation is restrained to day-time hours, *i.e.*, in-between 5 am and 8 pm. The data from the full month of February 2018 were used for the offline calculation of the trip lengths matrix $\hat{\boldsymbol{L}}$.

Firstly, each GPS trajectory was transformed into our trip representation (*i.e., regional path*, *actual arrival time*, and *actual travel time*), along with the *trip id*. This approach corresponds to introducing a spatial down-sampling to replace the precise track information with the regional path feature.

To reproduce the temporal sparsity of mobile phone data, we artificially introduced temporal biases based on a given value of the average inter-event time, as from our bias model. The value of the average inter-event time depends on the population observed or on the type of data chosen: for example, NSD datasets will display shorter inter-event times than CDR and LBSN. Therefore, we generated, for each day of data, five different down-sampled datasets, one per inter-event time value. The selected average inter-event time values were 4, 8, 12, 16, and 20 minutes, to cover a large range of average communication rates. The introduction of the bias was performed on a per-trip basis, by sampling the departure and arrival biases according to the probability density function defined via Eq. 3.12 using the average inter-event time values above. The actual travel time of each trip is thus biased with the sum of the sampled departure and arrival biases to obtain the observed travel time. Finally, the biased trips were grouped by regional paths and 15-minutes slots, averaging the observed travel times in each of the resulting groups thus obtaining the $\bar{T}_P^t$ values. By using our methodology as from Eq. 3.25, we obtained a speed profile in kilometres per hour, per region, and per 15-minutes slots for each day of the evaluation.

To evaluate the reliability of the results, we compared the estimated regional speed profile to the spatial mean speed V_r^t in region r over t. The latter was computed from the original GPS trajectories as the ratio of the total travelled distance TTD_r^t and the total travel time

Avg IET	MAE (km/h)	RMSAE (km/h)	MAPE (%)	RMSAPE (%)
4	7.894771	18.576767	10.773218	21.648403
8	12.970168	31.347989	16.233049	33.484246
12	16.918677	41.133133	20.600466	42.693440
16	19.920772	48.522253	24.008746	49.748655
20	22.204453	54.189049	26.608480	55.214964

Table 3.1: Average speed errors for each mean inter-event time (without de-biasing).

	MAE (km/h)	RMSAE (km/h)	MAPE (%)	RMSAPE (%)
Day 1	4.911907	7.174379	12.124498	15.264945
Day 2	5.057217	7.564531	12.024699	15.306312
Day 3	5.067316	7.505059	12.617445	15.839542
Day 4	4.848885	7.216612	13.006559	18.000039
Day 5	4.678821	6.591109	12.331655	15.714796

Table 3.2: Daily speed errors with de-biasing in the worst-case scenario (20-minutes IET)

TTT_r^t in region r during t, as from Eq. 3.2.

Tab. 3.1 reports the performance metrics (*i.e.,* Mean Absolute Error (MAE), Root Mean Absolute Error (RMSAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Absolute Percentage Error (RMSAPE)) for each day as a function of the average inter-event time, prior to applying the de-biasing procedure. The errors are quite high and increase significantly with the average inter-event time, highlighting the importance of the de-biasing method. This table can be contrasted with the results in Tab. 3.2, where the worst-case scenario, corresponding to an average inter-event time of 20 minutes, is analysed after applying our bias removal procedure. For all metrics and all days, there is a noticeable improvement in performance indicators due to the de-biasing procedure. The errors remain below a 20% limit when considering the daily RMSAPE, which is a significant improvement compared to the average value of 55.2% observed in Tab.3.1.

Fig. 3.3 displays the smoothed speed estimation results for the five different average inter-event time values, again after bias removal. The speed curves for different average inter-event time values almost overlap, confirming the importance and effectiveness of our bias removal approach. The results also show that the framework is generally capable of reproducing speed trends and dynamics (*i.e.,* good match with the blue curves in the figure), with particularly good results for many of the considered urban areas and almost all of the ring-road regions.

The less satisfactory results for some regions could be partially explained by the low quantity or quality of the original GPS data in those areas. This issue may be related to the presence of tunnels or a low penetration rate of the data provider in those areas, which inevitably reduces the representativeness of the average travel times, despite the artificial expansion of the dataset. In situations where the input data do not adequately describe traffic conditions, the bias removal mechanism is inevitably insufficient for producing accurate speed estimations. This highlights the need for more advanced processing solutions, such as those that will be described in Sec. 3.2. It is also worth noting that validating the obtained results on real NSD and CDR datasets remains a task for future work.

3.2 Trajectory Inference at Scale with Network Signalling Data

To address the challenge of accurately reconstructing mobility trajectories from real-world, massive individual NSD and fully exploit their potential for urban mobility, we proposed a framework named *TRANSIT*, extensively described in [151].

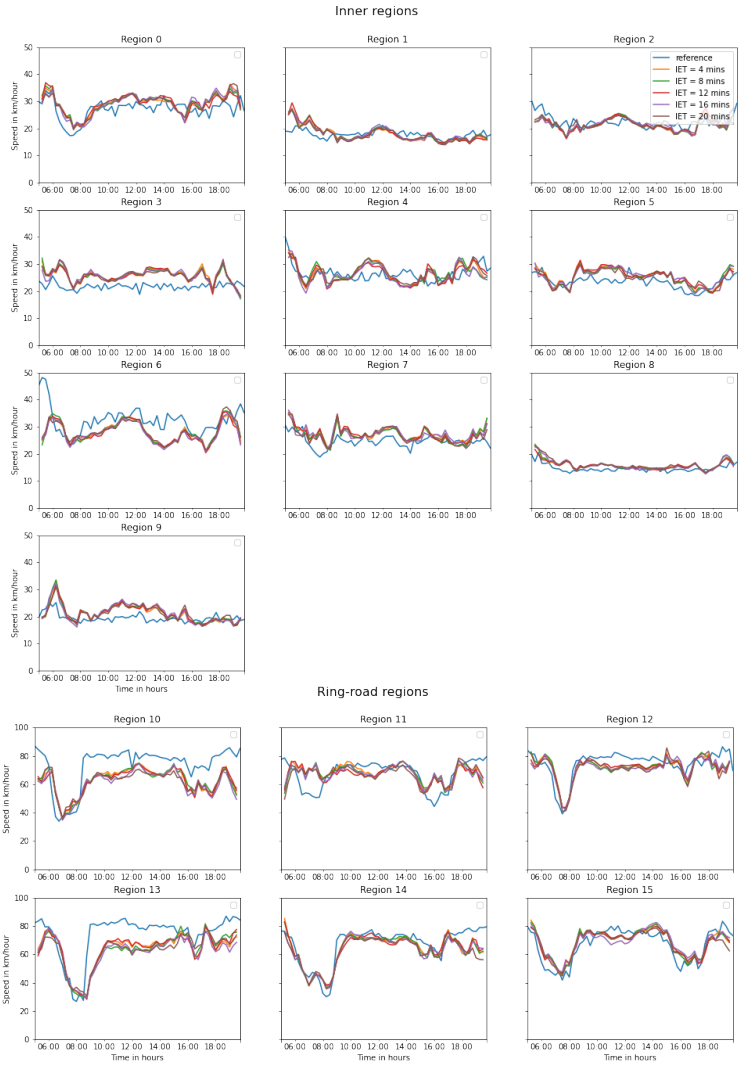


Figure 3.3: Speed dynamics after average bias removal. Ground truth speeds from the raw GPS traces are reported as blue curves.

TRANSIT was conceived to explicitly leverage the high sampling rate of NSD and the repetitive nature of human mobility, by building on interesting approaches from the literature recently proposed to overcome location-related limitations of CDR and CDR+ datasets: the *Detect, Expand, Check and REmove* (DECRE) [152] and the *Cumulative Weighted Moving Average* (CWMA) [153, 154] frameworks.

DECRE is built on the principle of removing oscillations from CDR in order to reduce spatial uncertainty for enhanced human mobility modelling. However, although such an approach could effectively improve spatial accuracy by removing the noise deriving from the oscillations, the resulting trajectory is still bounded to the original location information from the cellular network (*i.e.,* antennas coordinates), thus exhibiting large spatial uncertainty. As for CWMA, the approach leverages oscillations to infer with increased accuracy user locations. However, CWMA does not take into account the existence of high regularity in human movements, and consequently in mobile phone events, which TRANSIT exploits to enhance the spatio-temporal accuracy of the inferred mobility trajectories.

TRANSIT receives as input the set of NSD events of a mobile device i[10], denoted by

[10] As in Sec. 2.2, we make the simplifying assumption in the following that a device corresponds to a given

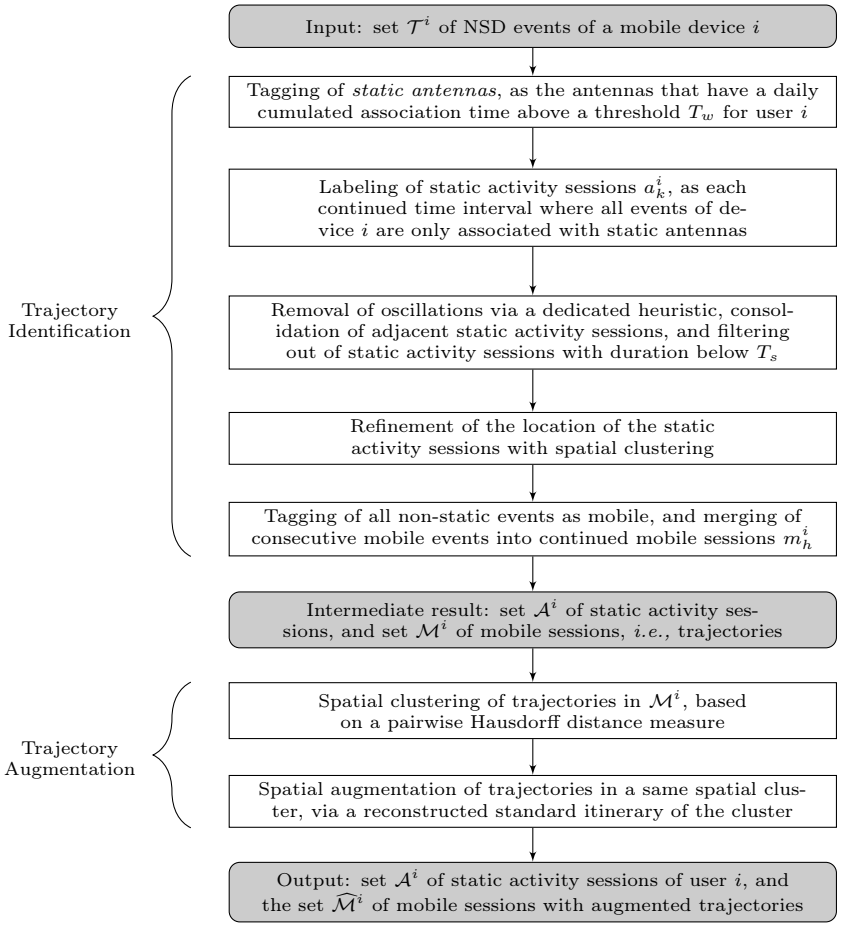


Figure 3.4: Flowchart of TRANSIT.

$\mathcal{T}^i = \{e_1^i, \ldots, e_n^i, \ldots, e_{N_i}^i\}$, where e_n^i is the n^{th} NSD event recorded for device i. Each NSD event is the result of a communication activity between a mobile device and a base station antenna of the telecommunication network, across all 2G, 3G and 4G technologies; it is defined as a tuple $e_n^i = (c_n^i, t_n^i)$, where c_n^i is the antenna at location l_n^i that handled the network event, and t_n^i is the timestamp of the instant at which the event was recorded. The NSD events in a mobile phone trajectory $\mathcal{T}^i$ are ordered by their timestamps t_n^i, and N_i denotes the number of events for device i. Then, TRANSIT processes $\mathcal{T}^i$ in two successive phases:

- *Trajectory identification.* The framework labels each NSD event $e_n^i \in \mathcal{T}^i$ as either static, if the user i is deemed to be engaged in an activity at a same location at the event time t_n^i, or mobile, if i is performing a movement at t_n^i. The labelling factually allows telling apart the continuous time intervals during which an individual is moving or not, and building a set $\mathcal{A}^i$ of *static activity sessions* and a set $\mathcal{M}^i$ of *mobile sessions*. As a result, the set $\mathcal{M}^i$ also identifies all the *trajectories*, *i.e.*, continued sequences of movement in time, of user i.

- *Trajectory augmentation.* The framework enhances the trajectories associated to mobile sessions in $\mathcal{M}^i$, by exploiting the fact that the same individual typically performs many trips between two given locations over time, generally following very similar paths. This creates redundancy in the mobility information that can be used to in-

individual.

crease the spatio-temporal accuracy of the trajectories. The resulting set of mobile sessions possibly augmented trajectories is denoted as $\widehat{\mathcal{M}}^i$.

Ultimately, the output of TRANSIT are the set $\mathcal{A}^i$ of static activity sessions of user i, and the set $\widehat{\mathcal{M}}^i$ of mobile sessions with augmented trajectories. Fig. 3.4 presents a flowchart of the stages of TRANSIT.

3.2.1 Trajectory identification

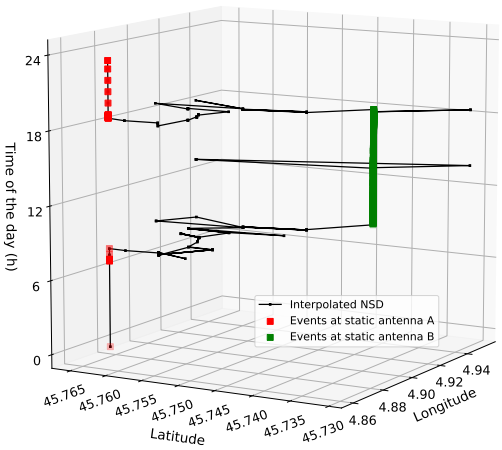
As anticipated, the trajectory segmentation step is applied to the individual set of NSD events $\mathcal{T}^i$ recorded for device i, and returns a subset of $\mathcal{T}^i$ where each event is labelled as static or mobile and detected oscillations are removed.

Fig. 3.5 illustrates the process of trajectory identification using TRANSIT. The interpolation of NSD events is portrayed as the black solid line. Fig. 3.5(a) refers to static antennas with daily accumulated association time above T_w. Fig. 3.5(b) identifies static activity sessions as obtained from consecutive sequences of static antennas only, and detected oscillations. Fig. 3.5(c) exhibits the final static activity sessions upon removal of oscillations, as well as the consequent detected mobile sessions.

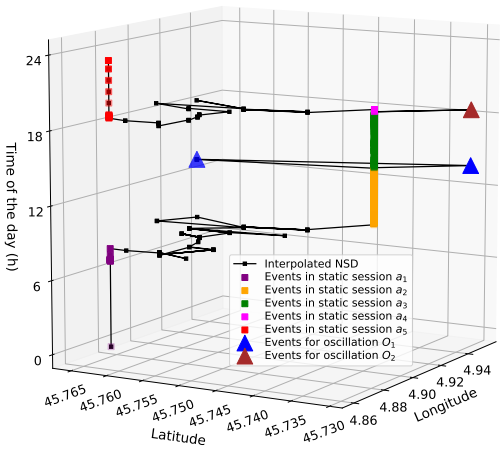
We start by assuming that the time spent by user i at the antenna c_n^i associated to event e_n^i is $t_{n+1}^i - t_n^i$, *i.e.,* the temporal span to the subsequent event e_{n+1}^i. Given the high temporal resolution of NSD, this simple approach already provides a very good estimation of the time the user is associated to a given antenna, at a low computational cost. Then, a preliminary labelling is performed to trim down candidate static events. To this end, we calculate the cumulated time spent by user i at each antenna c_n^i, on a daily basis. As devices stay connected to a limited set of antennas while still, we expect such antennas to yield a non-negligible cumulated time during the target day. We thus tag as *static antennas* for user i those antennas with a daily cumulated time above a threshold T_w. In our experiments, we set T_w to 20 minutes, which falls within the range of commonly accepted values for the typical minimum duration of a significant activity carried out by an individual at a same location [131, 155], and is employed also with high-frequency longitudinal (*e.g.,* GPS) data [156]. An example is provided in Fig. 3.5(a).

A continued time interval where all events of device i are only associated with static antennas is then denoted as a static activity session a_k^i. Typically, during one day, a user can have several static sessions, and each can be composed of one or multiple antennas. The set of all such sessions across the whole observation period is $\mathcal{A}^i = \{a_1^i, \ldots, a_{K_i}^i\}$. It is worth highlighting here an important aspect concerning the spatial extent of the static activities detected by TRANSIT: while the approach proposed in Sec. 2.2 for OD matrix estimation detects stationary activities at the level of EDR sectors, TRANSIT detects static sessions at the level of the coverage area of one (or small groups of) antenna(s), which, in urban areas, corresponds to an extent from one to three orders of magnitudes smaller than the average area covered by EDR sectors. This represents a significant advancement in terms of the spatial resolution adopted with TRANSIT. This also represents an improvement with respect to the mean speed estimation framework described in Sec. 3.1. That framework relies on a regional partitioning of the road network, where regions are defined by homogeneous traffic analysis zones. These zones typically cover a significantly larger spatial extent than the coverage area of antennas, especially in dense urban environments.

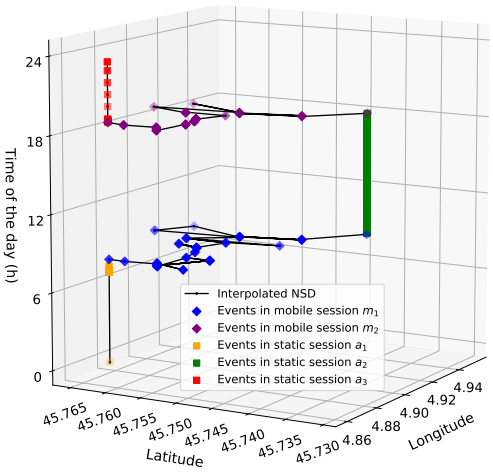
TRANSIT includes an additional step, addressing the removal of oscillations from the mobile phone traces, to enhance the estimation of the duration and location of the detected static sessions. After the stage above, only part of the antennas are labelled. Unlabelled antennas are either encountered during movements, or the result of oscillations that are known to characterise mobile device association to the radio access infrastructure [156]. Oscillations can in fact affect both static and mobile users. In the former case, they can cause the separation of continuous static activities into different static sessions in $\mathcal{A}^i$ interleaved



(a) Identification of static antennas



(b) Identification of static activities



(c) Oscillation removal and identification of mobile sessions

Figure 3.5: Main steps of the trajectory identification via TRANSIT.

by non-static antennas. In order to address the issue, and remove oscillations from $\mathcal{A}^i$, TRANSIT adopts the following heuristic. If (i) two consecutive static sessions a^i_k and a^i_{k+1} present at least one common (static) antenna, and (ii) the number of unique antennas associated to events observed after a^i_k and before a^i_{k+1} is below a threshold N_o, we merge all the events in a^i_k and a^i_{k+1} into a new, single static session. The new sessions replaces the former pair in $\mathcal{A}^i$. An example of oscillation detection and static sessions before the merging process is shown in Fig 3.5(b).

The single events identified as oscillations in the previous stage are in fact removed from $\mathcal{T}^i$ entirely, so as to limit uninformative noise in the data. The revised static sessions in $\mathcal{A}^i$ are further filtered based on their total duration, and only those with a time span higher than a threshold T_s are retained. The value of T_s corresponds to the assumed minimum duration of a static activity, so that we do not include, *e.g.,* waiting periods at red traffic lights for pedestrian or vehicular trips, or dwell times at stops for bus trips. For the same reasons explained above in relation to threshold T_w, used to identify static antennas, the value of 20 minutes has been adopted for T_s as well.

TRANSIT also enforces consistency in the locations of events associated to static activity sessions, as follows. First, we compute the centroid of the locations l^i_n of all events in each session a^i_k; then, the well-known DBSCAN clustering algorithm[11] is run on the centroids of all $a^i_k \in \mathcal{A}^i$. This lets us group together all static sessions related to a same activity, and compute a consolidated location for the activity as the barycentre of all centroids in a same cluster. The locations l^i_n of all events in each session a^i_k are then replaced with the barycentre of the corresponding cluster. Note that the position of the static activity sessions that are labelled as outliers by the DBSCAN algorithm are left unchanged. An example of the resulting $\mathcal{A}^i$ is in Fig 3.5(c).

Finally, all events that have not been labelled as static are labelled as mobile. This directly identifies the mobile sessions m^i_h of user i, as the time-continuous sequences of mobile events; an important remark is that the two static events immediately preceding and following the mobile session are also integrated into m^i_h to get a more precise indication of the start and the end locations of the corresponding trip. As a result, the set of mobile sessions is $\mathcal{M}^i = \{m^i_1, \ldots, m^i_{H_i}\}$. Each m^i_h corresponds to one trajectory of user i identified by TRANSIT. An example with two detected mobile sessions is reported in Fig 3.5(c).

3.2.2 Trajectory Augmentation

The sequences of NSD events in $\mathcal{T}^i$ that correspond to the single trajectories m^i_h of user i are still affected by the limited spatial accuracy that characterises NSD events as explained in Sec. 1.1.2. In its second phase, TRANSIT thus aims at improving the geographical quality of the movement information. As anticipated, the framework relies on the regularity of human mobility; more precisely, we use the the information from multiple similar trajectories identified for a same user to mutually improve their accuracy.

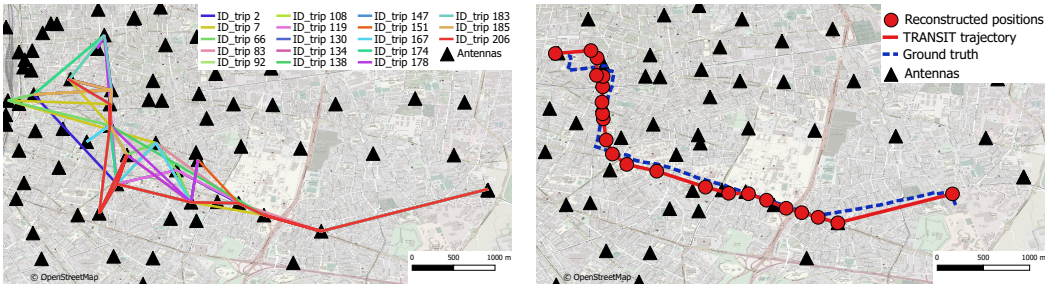
As a first step, a similarity measure is computed for all pairs of mobile session $m^i_h \in \mathcal{M}^i$. We employ the Hausdorff distance [157], which is defined as:

$$d_H(m^i_{h_1}, m^i_{h_2}) = \max\{D(m^i_{h_1}, m^i_{h_2}), D(m^i_{h_2}, m^i_{h_1})\}, \tag{3.27}$$

$$\text{where} \quad D(m^i_{h_1}, m^i_{h_2}) = \sup_{l^i_{n_1} \in m^i_{h_1}} \inf_{l^i_{n_2} \in m^i_{h_2}} d(l^i_{n_1}, l^i_{n_2}), \tag{3.28}$$

where $m^i_{h_1}$ and $m^i_{h_2}$ are the two mobile sessions to be compared and $d(\cdot,\cdot)$ is the geodesic distance between the two argument locations. This results in a matrix of pairwise distances between all mobile sessions of a same user i.

[11] Details on the parameters used with the DBSCAN algorithm for static session clustering can be found in [151].



(a) Set of trajectories of a same voluntary user clustered by DBSCAN

(b) Final trajectory, as inferred with TRANSIT

Figure 3.6: Trajectory reconstruction for a given pair of Origin-Destination path.

Then, DBSCAN is applied to the distance matrix, in order to group trajectories that have similar spatial geometries, and correspond to diverse trips of the user between the same two static activity locations. Fig. 3.6(a) shows an example of a set of mobile sessions, *i.e.,* trajectories, grouped together in the same cluster by DBSCAN, for the origin-destination activity locations in Fig. 1.1. Based on the result of DBSCAN, we can tell apart the mobile sessions in $\mathcal{M}^i$ into two subsets: (i) trajectories that fall into a cluster, *i.e.,* which refer to a path that is recurrent in the mobility of user i, and which we denote as the set $\mathcal{M}^i_R$; and, (ii) outlier trajectories that represent unique movements of i, which are grouped in set $\mathcal{M}^i_O = \mathcal{M}^i \setminus \mathcal{M}^i_R$.

For trajectories in $\mathcal{M}^i_R$, TRANSIT operates a spatial augmentation, as follows. First, the average duration is computed for all trajectories assigned to the same spatial cluster by DBSCAN above; this corresponds to the expected time that user i takes to travel between the same origin-destination activity locations. The time information is used to filter out trajectories whose duration deviates from the median by 50% or more: these mobile sessions are considered not representative of the routine mobility patterns along the target path. The retained trajectories in the same cluster are then temporally scaled (*i.e.,* stretched or compressed) in time so as to match the average travel duration for the cluster. Finally, the scaled trajectories are temporarily binned according to a fixed time period of one minute, and the spatial coordinates of all different events that fall in the same time bin are averaged.

The previous steps lead to a set of positions, one per minute, which represent the reconstructed itinerary. If there is no event within a particular time slot, the resulting enhanced trajectory will have missing positions. All trajectories in the cluster are then matched to the reconstructed one and become thus identical in the space dimension. However, they are re-conducted to their original duration (*i.e.,* via compression or stretching) so as to keep them faithful to their recorded travel time in the NSD.

As a result, each original mobile session in $\mathcal{M}^i_R$ is replaced by a set of reconstructed positions without any temporal deformation and is enriched with information derived from multiple similar trajectories travelled by the same user. This set of enhanced mobile sessions is referred to as $\widehat{\mathcal{M}^i_R}$. Fig. 3.6(b) shows the final spatial trajectory inferred from the cluster in Fig. 3.6(a). Trajectories in $\mathcal{M}^i_O$ stay instead unchanged, corresponding to those obtained from the simple interpolation of NSD data. The final set of mobile sessions is $\widehat{\mathcal{M}^i} = \widehat{\mathcal{M}^i_R} \cup \mathcal{M}^i_O$.

3.2.3 Performance Evaluation

The performance of TRANSIT has been evaluated by using ground-truth information on the trajectories of a small set of four volunteers. More precisely, we collected a total of 310 high-resolution trajectories using a high-frequency GPS logger running on their smartphones; also,

we recovered the NSD data generated by the volunteers' devices (which used Orange as their network provider) during the same observation period. We stress the fact that, although reduced in size, this validation dataset is among the few allowing a direct comparison of NSD and GPS data. The details of the validation process are reported in the reference paper [151], jointly with extensive sensitive analyses of the selected parameters as well as a thorough investigation of the impact of mobility recurrence on the accuracy of the results.

Here, we report the key figures obtained with TRANSIT on the validation dataset, relative to the two identified state-of-the-art solutions (*i.e.,* DECRE and CWMA). We used the classical *precision*, *recall* and *F1* metrics to evaluate the performance of the trajectory segmentation approaches. Formally:

$$\text{Precision} = \frac{TP}{(TP+FP)}, \quad \text{Recall} = \frac{TP}{(TP+FN)} \quad \text{and} \quad \text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{(\text{Precision} + \text{Recall})} \tag{3.29}$$

where *(i)* the number of true positives TP is the number of NSD events labelled as static when the user is also labelled as static in GPS data; *(ii)* the number of false positives FP represents the number of NSD events labelled as static while the user is in fact mobile according to the ground truth; *(iii)* the number of false negatives FN maps to the number of NSD events labelled as mobile while the user is static in the GPS data. Specifically, TRANSIT yields a 5% relative improvement, in terms of the *F1* metric, with respect to CWMA in the trajectory identification task (we remind that DECRE does not include any classification procedure), with nearly perfect recall of static sessions.

More importantly, concerning trajectory enhancement, TRANSIT attains an average geographical error as low as 190 meters, bringing it down to 130 meters for some users on average, with global improvements in the order of 30% and 41% with respect to CWMA and DECRE, respectively, when used on the same NSD datasets. The improvement raises to 60% when CWMA is used on the CDR+ version of the dataset, and 72% when DECRE is used on the CDR one. These results were obtained using the average of two distance metrics computed on a per-trajectory and per-user basis as follows:

$$D_{GPS} = \frac{1}{|m_{GPS}|} \sum_{e_{n'} \in m_{GPS}} \min_{e_n \in m_{NSD}} d(l_{n'}, l_n) \tag{3.30}$$

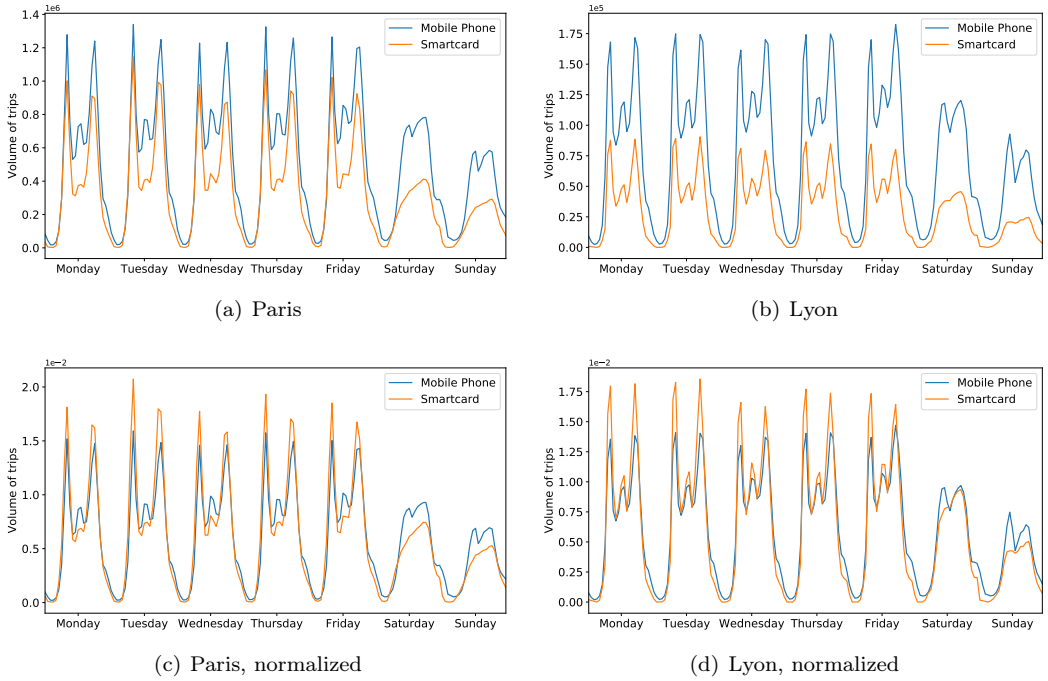
$$D_{NSD} = \frac{1}{|m_{NSD}|} \sum_{e_n \in m_{NSD}} \min_{e_{n'} \in m_{GPS}} d(l_n, l_{n'}) \tag{3.31}$$

where m_{GPS} and m_{NSD} are two corresponding trajectories inferred for a given user from GPS and mobile network data, respectively. The operator $|\cdot|$ denotes the cardinality of the argument set, *i.e.,* the number of samples in the case of a trajectory, and $d(\cdot,\cdot)$ the geodesic distance. We use both metrics as they are complementary: while D_{GPS} is representative of the error observed for continuously tracked user, D_{NSD} measures the error specific to events recorded by the mobile phone network.

The takeaway is that the relative improvement of TRANSIT does not simply derive from the increased temporal resolution of NSD, but also by its capacity to enhance recurrent trajectories via historical data.

3.2.4 Large-Scale Applications of TRANSIT

While the validation results were related to a reduced number of users, the interest of TRANSIT reveals at city-wide scales, where it can enable a number of mobility-related applications. In the following, we thus focus on four case studies of TRANSIT related to urban mobility that fully leverage the large-scale NSD datasets $\mathcal{D}_P$ and $\mathcal{D}_L$ provided by Orange Innovation for the cities of Paris and Lyon, respectively, and introduced in Sec. 1.1.2. We stress the fact that such results have been obtained by means of specific optimisations



(a) Paris

(b) Lyon

(c) Paris, normalized

(d) Lyon, normalized

Figure 3.7: Average weekly profiles of the number of concurrent trips in (a) Paris and (b) Lyon, as inferred from TRANSIT and smart card data. Normalised versions of the same profiles are presented in (c) and (d).

required in order to process the three months of NSD from the large-scale datasets $\mathcal{D}_P$ and $\mathcal{D}_L$. In particular, TRANSIT has been implemented in Spark by exploiting the Map-Reduce distributed programming paradigm.

Urban Mobility and Public Transport

By counting the number of concurrent active trips inferred via TRANSIT over time, we were able to reconstruct accurate temporal profiles of the travel demand in urban regions. For such profiles to be dimensionally correct, a re-scaling procedure was needed to account for the penetration rate of the technology (close to 100% in developed countries like France) and the market share of Orange (at 37% over the French territory). The resulting average weekly demand profiles computed in Paris and Lyon are depicted in blue in Fig. 3.7(a) and Fig. 3.7(b), respectively. Our estimates are that around 1,300,000 individual trips occur at the same time in Paris during commuting peaks, while the figure is at 180,000 for Lyon.

We compare the profiles obtained with TRANSIT with equivalent ones from smart card data, which capture mobility via public transportation systems. For Paris, data were provided by the transportation company Ile-de-France-Mobilité. Concerning Lyon, data were shared by the transportation company Keolis-Lyon. For both cities, public transport data were provided in the same period of the year of NSD, and all smart-card transactions were anonymised in the form of aggregate measures at the scale of the whole agglomeration.

Also in this case, a re-scaling was required: while the TRANSIT trajectories refer to the resident population, the smart card data include both residents and non-residents. In order to make the numbers comparable, we applied a scaling factor of 0.81 to the smart card temporal profile; the factor has been calculated from the raw network signalling data, by computing the average instantaneous fraction of resident subscribers present in the target cities, over the total number of observed users. The weekly profiles from smart cards are su-

perposed to the TRANSIT-inferred ones, as the orange curves in Fig. 3.7(a) and Fig. 3.7(b).

The comparison of the profiles reveals interesting facets of mobility in Paris and Lyon. Clearly, the volume of trips identified by TRANSIT is higher than that reconstructed with smart card data: NSD allows monitoring virtually all transport modes, including those beyond public means, *e.g.,* private vehicles, biking, or walking. This lets us quantify which proportion of trips is performed by underground, buses or tramways, and which using personal means. We find that a significant fraction of trips is performed using public transport in both cities: we estimate the percentages of movements captured by smart card data to 66% and 39% of the total, in Paris and Lyon, respectively. The difference between these values is explained by the more developed multi-modal transit network available in Paris, as required to support mass mobility in such a large metropolis.

In addition, we can investigate the temporal incidence of public transport by looking at versions of the same profiles that are normalised so that the integral of all curves is one. Fig. 3.7(c) and Fig. 3.7(d) show the result. This perspective lets us appreciate how in Paris public transport is especially important during commuting hours, but relatively less used during the lunch break or weekends. A slightly different pattern emerges in Lyon, where public transport is also intensely used around midday, but has a lower incidence during evenings and Sunday mornings. We highlight that obtaining this type of insight is hardly achievable by solely relying on surveying, which demonstrates the value of NSD and a method like TRANSIT that can exploit them.

As a final remark, we highlight that the results in Fig. 3.7 can also be considered as a partial validation of the trajectories inference mechanism performed by TRANSIT in large-scale settings. Indeed, the near-perfect match of the timing of commuting peaks or overnight low mobility among curves proves the capability of our trajectory segmentation approach to identify trips that are very consistent with data collected in the field over time. It is also worth underlining that, differently from the results reported in Sec. 2.2.2 and Sec. 3.1, no de-biasing procedure was applied in relation to the start time of the mobility session or the land use. The fact that the profiles from the two different sources align so well, without any de-biasing mechanism, further proves the relevance of 4G communication events in the dataset and the fundamental contribution of the processing taking place in the trajectory identification and enhancement phases of TRANSIT.

Popular Paths of Commuting Trips

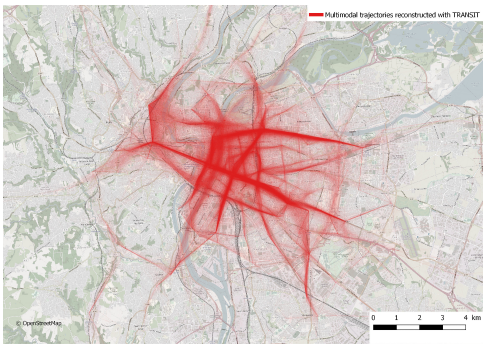
As a second application, we focused on inferring popular commuting trips within a city. The knowledge of such trips is an extremely precious source of information for transport authorities and city planners as: *(i)* they represent the largest share of the daily urban traffic demand of a city; *(ii)* they identify the typical commuting behaviours of travellers which regularly stress the transport network infrastructure, especially during peak hours; *(iii)* they are hard-to-quantify and characterise at city-scale because of the absence of dedicated sensors or probes that can precisely capture the multi-modal, diverse and time-varying nature of such trips.

By applying our framework to the large-scale datasets $\mathcal{D}_P$ and $\mathcal{D}_L$, we associated to each user i of the two analysed cities a set of trips $\widehat{\mathcal{M}}^i$ for the whole period of 3 months. As explained above, $\widehat{\mathcal{M}}^i$ can be divided into two subsets: $\widehat{\mathcal{M}}^i_r$, a subset of recurrent trips enhanced by TRANSIT, and $\mathcal{M}^i_o$, a set of non-recurrent trips of user u. Considering that commuting trips are, by definition, recurrent, only subset $\widehat{\mathcal{M}}^i_r$ was retained in this analysis. Furthermore, to extract commuting trips from $\widehat{\mathcal{M}}^i_r$, we filtered only those trips associated with the two most popular locations of each user, under the constraint that at least 10 trips were present between these two locations. The underlying assumption was that the remaining set should mostly contain the two most popular trips performed by users in their daily routine, *i.e.,* home-to-work and work-to-home trips (commuting trips).

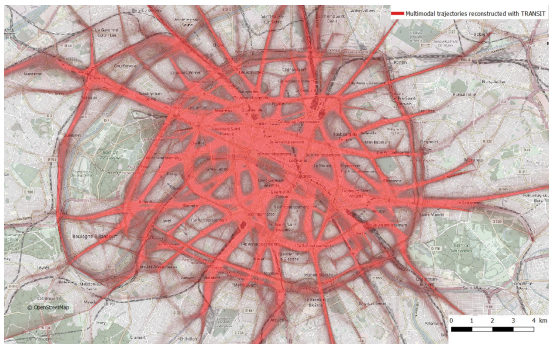
The spatial density (heatmap) of the reconstructed trips is represented in Fig. 3.8(a)

for Lyon and Fig. 3.8(b) for Paris. As a first consideration, the recurrent trips appear to have overall a good match with the multi-modal urban transportation network, graphically overlapped to the heatmap in Fig. 3.8(c) for Lyon and in Fig. 3.8(d) for Paris. A more in-depth inspection of the figures highlights that, for both cities, the subway network, the tramway lines and most important urban roads clearly show up among the commuting trips reconstructed via TRANSIT. In the case of Paris, NSD trips appear to have a near-perfect match to the underlying multi-modal transport network. The less evident match for the case of Lyon, especially characterising some peripheral roads (however present in the heatmap), can be explained by the lower number of available trips and the lower density of the cellular network of Lyon in these areas, compared to those from the capital city.

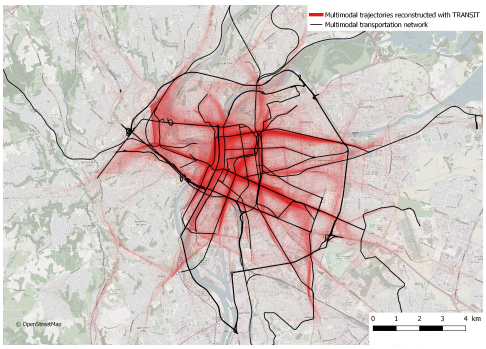
Of course, the fact that the majority of commuting trips maps to the public transportation network is not unexpected. However, TRANSIT opens the door to a detailed analysis of these trips: the obtained trips can be map matched to the different transportation lines and modes (see Sec. 3.3), showing their share of trips, on different days of the week and at different times of the day. Such information would be highly valuable for any public transportation company or municipality to perform, *e.g.,* informed planning and efficient resource deployment. Similarly, from a resilience perspective, highly congested routes or transit lines could be easily spotted, even in real-time settings, for network monitoring and vulnerability assessment.



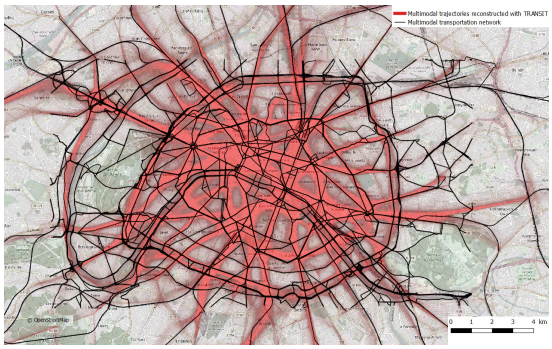
(a) Heatmap of trips ($D > 3km$) in Lyon



(b) Heatmap of trips ($D > 3km$) in Paris



(c) Heatmap of trips ($D > 3km$) in Lyon with the multi-modal transportation network of Lyon

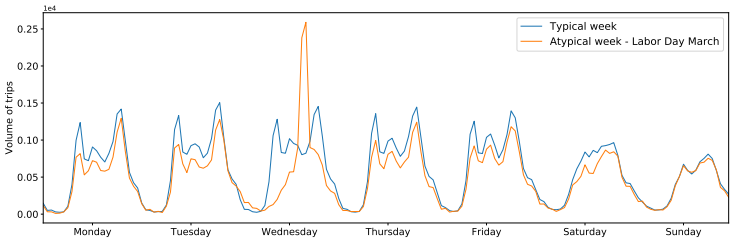


(d) Heatmap of trips ($D > 3km$) in Paris with the multi-modal transportation network of Paris

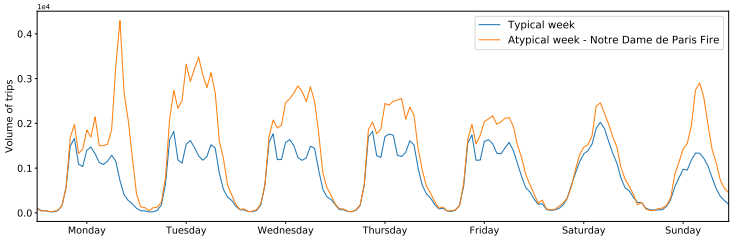
Figure 3.8: Heatmap of commuting trips in Lyon and Paris.

Human Mobility Analysis during Abnormal Events

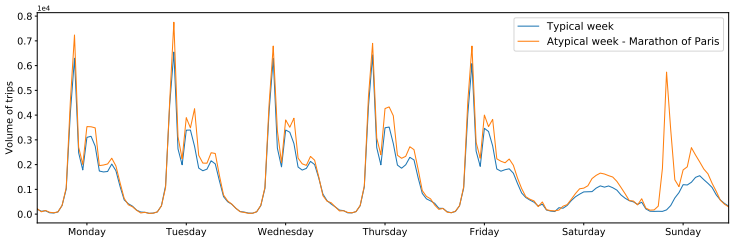
As a third application, we used TRANSIT to detect abnormal mobility situations that can occur in the city. In the perspective of resilient management of urban mobility such knowledge could prove to be essential to learn how people react to specific kinds of abnormal urban situations, such as accidents, sports events, and celebrations. For this, we segmented the city of Paris into a grid with squares of dimension 800m $\times$ 800m, with a temporal bin size of one hour. This spatio-temporal granularity makes it possible to analyse human mobility at a fine-grained scale. For each zone, we computed the *attraction demand profile*, which corresponds to the number of trips having as destination the studied zone at any given hourly time slot. These profiles were obtained by retaining such trips from the whole set of trajectories $\widehat{\mathcal{M}}^i$ computed via TRANSIT on $\mathcal{D}_P$ for each user i. This allowed us to build a typical weekly attraction profile for each zone and, at the same time, to distinguish abnormal patterns during certain events. We considered three such abnormal mobility situations, reported as examples below.



(a) Attraction of the zone Place d'Italie (Paris)



(b) Attraction of the zone Notre Dame (Paris)



(c) Attraction of the zone Arc de Triomphe (Paris)

Figure 3.9: Typical/atypical weekly temporal demand profile during atypical events.

First of all, on Wednesday, the 1st of May 2019, a bank holiday, the Labour Day march took place near Place d'Italie in Paris. Fig. 3.9(a) shows in blue the typical attraction profile of this zone and in red the attraction profile of the week that includes the demonstration. Whereas the attraction profile was similar to the typical profile for most days, we can see that, on Labour Day, the attraction of the studied zone presents an unexpectedly high peak after midday.

As a second event, we studied the fire of Notre Dame de Paris cathedral, on Monday the

15$^{\text{th}}$ of April 2019. Fig. 3.9(b) shows in blue the typical attraction profile of this zone, and in red the attraction profile of the week that includes the abnormal event. We can see a high peak in the attraction profile right after the beginning of the fire on Monday 15$^{\text{th}}$ (around 6:30 pm). Contrary to the previous example, this event also affected mobility the following days, when an attraction demand higher than usual is observed in the corresponding area. The attraction profile progressively decreases after the event, but an upsurge on Sunday, the Easter holiday, can be noticed, probably explainable by religious activities and nearby gatherings of tourists and worshippers visiting the area surrounding the cathedral after the fire on this special day.

Finally, we studied another special event, the Marathon of Paris, on Sunday the 14$^{\text{th}}$ of April 2019, with its start and end in the proximity of the Arc de Triomphe. A high peak on the attraction profile can be observed at the departure time of the marathon, at 9 am, as shown in Fig. 3.9(c). A second peak is observed a few hours later, more spread over time and lower in magnitude compared to the first one, corresponding to the marathon arrival.

These three examples are representative of the vast potential of TRANSIT towards building mobility profiles of the typical demand attracted by (or emitted from) a given zone, as well as detecting and characterising mobility patterns during abnormal or special events. These ideas were at the core of the real-time anomaly detection approach developed in the context of the ANR project DISCRET and discussed in [158].

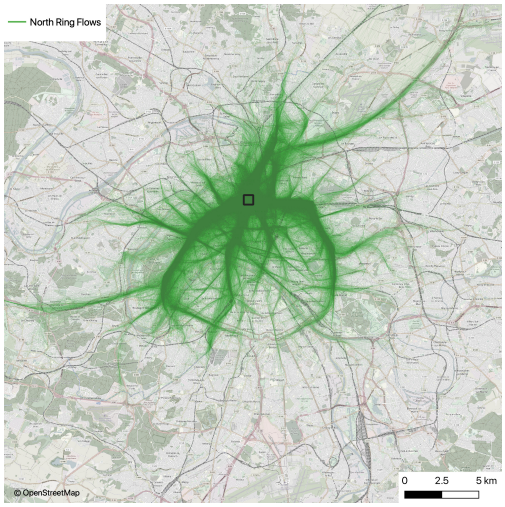
Ring Road Trajectory Analysis

As a fourth application, we leveraged TRANSIT to perform a fine-grained trajectory analysis focused on the Paris *périphérique* (ring road). The mobility flow on this urban highway is usually very high, often leading to heavy congestion, especially during peak hours. Transport authorities are traditionally very interested in the possibility of tracing and quantifying the flows of people moving along city major road axes. Such studies are necessary for urban planning purposes, infrastructure renewal and road maintenance, and can be extremely cost-demanding. They are normally based on travel diaries or GPS trajectory collection and generally end up capturing only a small sample of the flow actually traversing the major axis, with resulting limited accuracy. TRANSIT allows leveraging NSD to access a much larger and more representative sample of this specific population.

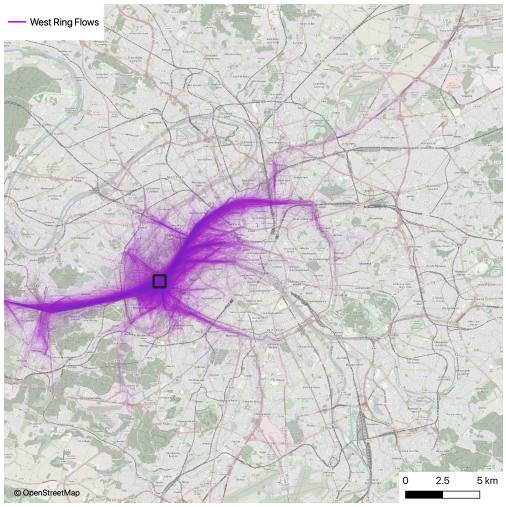
In our case study related to the Paris *périphérique*, we considered four different zones of interest: the east, west, north and south entries. The idea was to select a spatial zone and study all the trajectories passing by the respective zone. The enhanced trajectories $\widehat{\mathcal{M}}^i$ produced by TRANSIT on $\mathcal{D}_P$ allowed us to capture at scale the origin, the destination, and the paths taken by the users passing by the studied zone, the kind of information usually expected in the aforementioned studies. The result for the four zones of the *périphérique* (north, west, south and east) are reported in Fig. 3.10(a), Fig. 3.10(b), Fig. 3.10(c) and Fig. 3.10(d).

The obtained maps underline the major role of the *périphérique* in Paris, allowing people to travel across the city and reach any area of interest. Some interesting patterns can be distinguished as well. For example, the trips coming from the west side of the city show a strikingly different pattern than the three other maps, with trajectories that appear less scattered towards the centre of Paris and concentrated along a few major axis. This can be explained by the fact that the west side of Paris is the richest area of the city, with inhabitants who have a lifestyle involving shorter commuting trips. Moreover, the west side of Paris is also the area with the highest density of offices, including the *La Defense* and *Boulogne* neighbourhoods. This could explain why this area attracts a large amount of trips, ever from faraway zones.

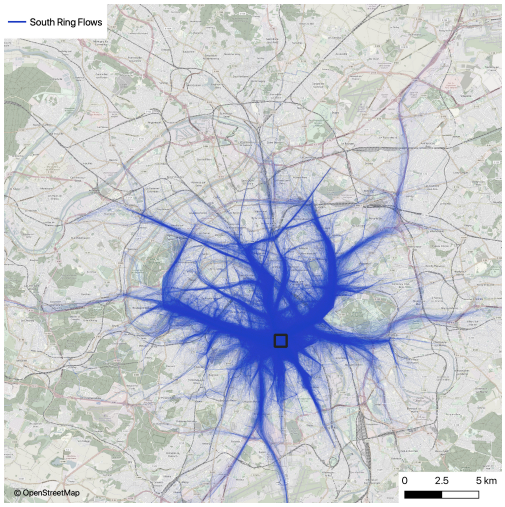
These results hint at the numerous perspectives brought by TRANSIT in the study of major road arteries. These include fine-grained temporal analysis, the detection of usage and attraction patterns, origin and destination profiling, *etc.*Ġenerally speaking, having access to



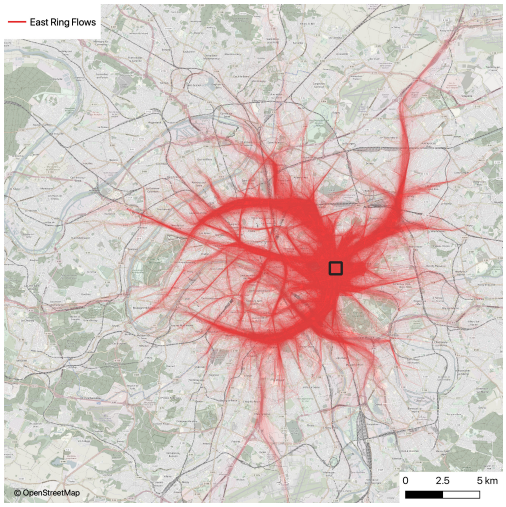
(a) North Ring Road



(b) West Ring Road



(c) South Ring Road



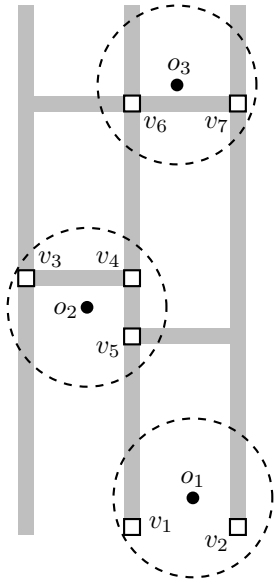
(d) East Ring Road

Figure 3.10: Heatmap of recurrent trips for the Paris ring road (the black square shows the catchment area)

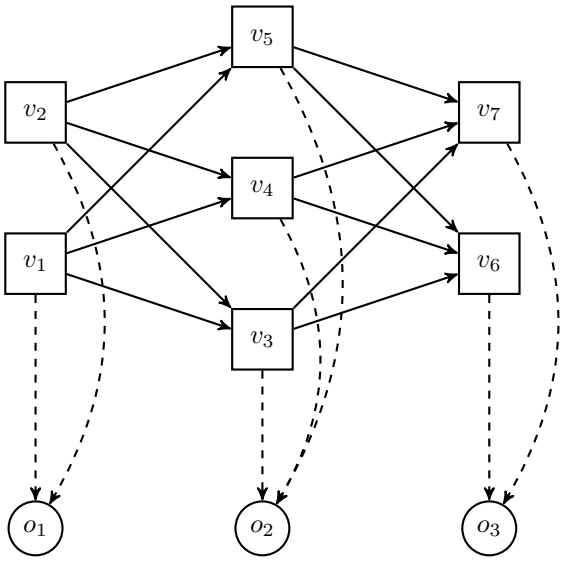
detailed human mobility trajectories at scale, such as those produced by TRANSIT, enables the in-depth study of any part of the transportation network.

3.3 Multi-modal Path Reconstruction via Hidden Markov Model

The high spatial accuracy of the outputs produced by TRANSIT naturally raises the research question of inferring the precise route taken by a mobile subscriber across the multi-modal transportation network of a city. In other words, the goal was to reconstruct the sequence of road segments, public transport stops or railroad paths followed by the observed user during a trip. This challenge, known as multi-modal map-matching of sparse spatio-temporal tracks, allows for achieving an extremely high level of detail in reconstructing



(a) Example of signalling trajectories (o_1, o_2 and o_3) and candidates nodes (v_1, v_2, $\cdots$ v_7)



(b) Hidden Markov Model. Solid arrows indicate transition probabilities, while dotted arrows represent emission probabilities. Pairs with null emission/transition probability have been filtered out

Figure 3.11: Illustration of Hidden Markov Model based map-matching.

mobility trajectories paving the way to precise multi-modal pattern detection and traffic variable estimation. In our specific case of NSD trajectories, we need to project the mobile phone trace, potentially improved through TRANSIT, onto a multi-layer graph used to represent the multi-modal transportation network of the target city. The problem is exemplified in Fig. 3.11(a) from [159]: an ordered sequence of timestamped points (o_1, o_2, o_3) corresponding to observations of a moving user has to be matched to a sequence of nodes, *e.g.,* (v_1, v_5, v_4, v_6), defining a path on the underlying graph-based representation of the transportation network. We addressed such research problem in [151, 160] by relying on Hidden Markov Model (HMM)-based map-matching.

HMM-based map-matching represents the state-of-the-art for noisy and sparse location data and, specifically, for mobile phone data traces [161]. A HMM can be defined as a 5-tuple $\langle S, O, I, T, R \rangle$, with $S = \{s_0, \ldots, s_{N-1}\}$ representing a finite set of states; $O = \{o_0, \ldots, o_{M-1}\}$ corresponding to a finite set of observations; I being the probability distribution of the initial states; T representing a set of transition probability. The probability to transit from hidden state s_m to hidden state s_n is denoted as $t(s_m, s_n)$. Finally, R is a set of emission probability. The probability to emit observation o_k from hidden state s_m is denoted as $r(s_m, o_k)$. The matrices I, T and R represent the main hyper-parameters of the HMM. Fig. 3.11(b) shows a possible HMM modelling of the toy example reported in Fig. 3.11(a).

In this context, we modelled the multi-modal transport network under study as a weighted directed multi-layer graph [56], $G = (V, E, L)$, with $L = \{l_q\}_{q=1}^{Q}$ the set of elementary layers, each representing a specific transport mode q (among a total of Q different options), V the set of nodes and E the set of edges. Each layer $l_q \in L$ contains a node subset $V_q \subset V$ and an edge one $E_q \subset E$. Each V_q corresponds to the set of road intersections, transit/train stops, *etc.*, associated to mode q. Each E_q represents the set of connections (road segments, public transport lines, *etc.*) between the nodes of the same transport mode q. Edges are directed according to the allowed direction of movement on the corresponding segment and weights can represent, *e.g.,* travel time, length, capacity or speed information associated to

the segment. In the general model, the edge set E also includes a set of inter-layer edges $E_{(l_{q_i}, l_{q_j})}$ that allows to connect pairs of layers l_{q_i}, l_{q_j}, *i.e.,* to switch between the two different transport modes q_i and q_j. In our study, we assumed graph G to be composed only of layers related to road, train and public transport, denoted as G_{road}, G_{train} and G_{pt} in the following. It is worth noting that, with respect to the general network modelling approach, TRANSIT allows the simplifying assumption to consider layers not connected to each other, *i.e.,* $E_{l_{q_i}, l_{q_j}} = \emptyset$. The rationale behind this assumption is the following: in order to move from one transport mode to another the user will need to spend a period of *immobility* at a given location (*e.g.,* in the proximity of a bus stop, *i.e.,* under the coverage of a limited subset of cellular antennas). If long enough, TRANSIT detects this waiting time as a static activity of the user. Therefore, a multi-modal trip over these three networks will actually be split in multiple sub-trips by TRANSIT, that can be each separately matched to the specific sub-graph. Inter-modal trips are therefore not possible across the three main sub-graphs G_{road}, G_{pt} and G_{train}, but can occur within the public transport network, *i.e.,* G_{pt}, which covers indeed three different transportation modes (*i.e.,* subway, tramway and bus) and is therefore a multi-layer graph itself.

In [160], we extended the TRANSIT framework with map-matching capabilities to further enhance the spatial accuracy of its output: map-matching is performed on the whole set of mobile sessions produced by TRANSIT, *i.e.,* $\widehat{\mathcal{M}^i}$, which includes both raw signalling trajectories ($\mathcal{M}^i_O$) and TRANSIT-enhanced recurrent trajectories ($\widehat{\mathcal{M}^i_R}$). Specifically, we used the following HMM formulation. Hidden states are modelled as the set of vertices (nodes) of the generic transportation sub-network G_j. Emissions are modelled as the unique set of x-y coordinates in $\widehat{\mathcal{M}^i}$. This set is composed of antennas coordinates from cellular network in $\mathcal{M}^i_O$ and reconstructed positions in $\widehat{\mathcal{M}^i_R}$. Ultimately, given a sequence of observations, *i.e.,* sequence of antennas for $\mathcal{M}^i_O$ and reconstructed positions for $\widehat{\mathcal{M}^i_R}$, the HMM is used to retrieve the most likely sequence of hidden states, *i.e.,* sequence of nodes on the transportation sub-network G_j.

In the related literature, a similar problem has been solved in [162, 163, 161]. However, these approaches have two main drawbacks. They leverage supervised machine-learning solutions that require a large amount of labelled cellular trajectories for training the parameters of the models. Ground-truth mode-labelled trajectories are very hard to obtain, especially when dealing with highly dynamic and complex environments, such as urban areas. Moreover, most of the approaches match the cellular trajectories only to road networks, without considering other sub-networks corresponding to alternative transportation modes, such as tramway, subway, bus, *etc.*

Among the very few exceptions, it is worth mentioning the methodology proposed by Asgari *et al.* [164]. Their solution, namely CT-Mapper, relies on an unsupervised HMM model, which aims at mapping sparse cellular trajectories[12] to a multi-layer transportation network. Similarly, in our previous work [151], we studied the application of unsupervised HMM-based map-matching for solving the same problem in more realistic settings, characterised by the usage of real-world raw NSD and the analysis of a dense urban case study.

Our approach builds on such previous work and solves the HMM-based map-matching problem according to a two-steps map-matching procedure. The first phase consists in an optimised Viterbi algorithm [165]. The inputs of the Viterbi procedure are the following: the generic transportation sub-network modelled as a (multi-layer) graph G_j, the possible hidden states (set of the nodes of G_j), the emissions (the unique set of x-y coordinates in $\widehat{\mathcal{M}^i}$), the HMM hyper-parameters (matrices I, T and R) and the input trajectory from $\widehat{\mathcal{M}^i}$. By calculating all possible paths given the input trajectory, the output of the Viterbi algorithm is the most likely sequence of graph nodes, one for each time instant in the input. For real-time application, due to a large number of states and emissions, execution time is

[12] In their work, Asgari *et al.* used a variant of CDR data, characterised by localisation information produced at a frequency, much higher than that observed with traditional CDR.

critical. To improve performance, similarly to [163], we implemented an optimised version of the Viterbi algorithm that eliminates all multiplications by zero thus reducing the search space by keeping only states that can be emitted from each observable state. After inferring the most likely states sequence using the optimized Viterbi implementation discussed above, the output sequence of hidden states (nodes on a given sub-network G_j) does not necessarily produce a connected path on the specific transport sub-network. Therefore, as the second phase of the map-matching procedure, the final trajectory is further completed by applying a traditional weighted shortest path detection algorithm on the underlying transportation graph between any two consecutive nodes of the most likely sequence. The final completed sequence of nodes on sub-network G_j represents the map-matched trajectory for the processed trace from $\widehat{\mathcal{M}}^i$ for user i.

3.3.1 Empirical Evaluation in the Lyon Case Study

In the following, we briefly describe the parametrisation of the HMM, as well as the main results of the proposed solution that has been applied to the $\mathcal{D}_L$ dataset of the city of Lyon. More information about the pre-processing steps, parameter choices and detailed results can be found in [160].

Initial Probability

All the nodes of the transportation network are initially equally assigned with a probability of $1/N$ with N representing the total number of nodes in the transportation (sub-)network:

$$\pi(s_m) = \frac{1}{N} \tag{3.32}$$

Transition Probability

The transition probability corresponds to the probability that a mobile phone user moves on the underlying transportation (sub-)network from hidden state s_m at time $t-1$ to hidden state s_n at time t. As in [166], the transition probability is assumed to depend on the travel time over an edge. For the public transport and railway sub-networks, the travel time of each edge is calculated by multiplying the reference free-flow speed associated to each mode[13] and the edge length[14]. Additionally, for public transport, cross-layers edges connecting the different lines and modes are associated to a typical waiting time of 5 minutes.

Finally, the transition probability $t(s_m, s_n)$ between the generic pair of nodes s_m and s_n is defined to be exponentially decreasing according to the travel-time weighted shortest path between the two nodes s_m and s_n. Formally:

$$t(s_m, s_n) = e^{-\beta \cdot tt_{s_m,s_n}}, \quad tt_{s_m,s_n} = \sum_{\forall (s_u,s_v) \in SP_{mn}} tt_{s_u,s_v} \tag{3.33}$$

where (s_u, s_v) is the generic edge on the travel-time weighted shortest path SP_{mn} connecting the two nodes s_m and s_n in sub-graph G_j, computed via the Dijkstra algorithm. The length of the weighted shortest path SP_{mn} corresponds to the sum of the travel time over each edge (s_u, s_v) belonging to SP_{mn}, while tt_{s_u,s_v} denotes the travel time between each pair of nodes s_u and s_v. Finally, β is a damping factor controlling the impact of the travel time on the transition probability.

[13] Speeds on the road network depend on the OSM type of route, ranging from 30 km/h to 90km/h. For the subway, the tramway and the bus the reference speed has been estimated respectively to 30 km/h, 15 km/h and 15 km/h.

[14] The length of an edge has been computed as the geodesic distance between the two adjacent nodes of the edge.

Emission Probability

Similarly to [167], we modelled the emission probability as a Gaussian spatial noise centred around the hidden state s_m and an empirically estimated standard deviation of the distance error between hidden states and observations:

$$r(s_m, o_k) = \frac{1}{\sqrt{2\pi}\alpha} e^{-0.5\left(\frac{d_{s_m,o_k}}{\alpha}\right)^2} \tag{3.34}$$

where d_{s_m,o_k} is the geodesic distance between the generic observation o_k and the generic node s_m, while α is the standard deviation of a Gaussian random variable associated to the error distance between the reconstructed and the real position of the mobile. Both α from Eq. 3.34 and β from Eq. 3.33 are hyper-parameters that need to be properly estimated via a sensitivity analysis.

Results

To evaluate the performance of the proposed HMM map-mathing solution, we used a refined version of the same small dataset introduced in Sec. 3.2.3 for the performance evaluation of TRANSIT. This dataset was also used to select the values of the α and β parameters of Eq. 3.33 and Eq. 3.34 via a sensitivity analysis fully detailed in [160]. Concerning the GPS trajectories, we manually labelled the transport mode of all trajectories by associating one sub-graph G_j of G for each trajectory. The final ground-truth data contain 111 trajectories related to public transport and 72 to car, for a total of 183 trajectories. TRANSIT was applied to the NSD data available for the same set of volunteers to retrieve the mobile sessions with augmented trajectories. The latter were used as the input of our map-matching approach.

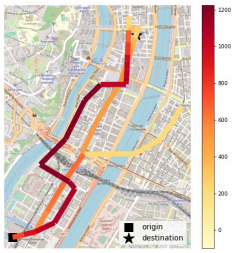
The results of the map-matching procedure were produced with and without prior knowledge on the transportation mode. In the case of map-matching performed without any prior knowledge on transportation mode, trajectories were matched to each of the considered sub-graphs of G. The final reconstructed output was chosen as the one with the highest likelihood from the Viterbi algorithm. In the other case, we considered that one could obtain a rough estimation of the transport mode, such as the average speed of the user from the observed NSD trajectory. With this information, and by associating a reference speed to each sub-graph, it becomes possible to identity the most likely candidate sub-graph for the reconstruction of the path for the given NSD trajectory.

To evaluate the map-matching performance, we used, among others, the distance metric defined in Eq. 3.31 to measure the geographical error, denoted as G_e in the following, and the matching rate, denoted as MR and defined as the percentage of correctly map-matched edges by our approach. Formally:

$$MR = \frac{TP}{TP + FN + FP} \tag{3.35}$$

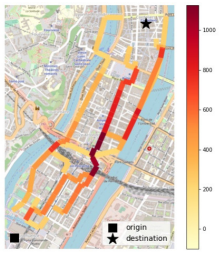
where: (i) the number of true positives TP is the number of edges in common between the ground-truth GPS and NSD map-matched trajectories; (ii) the number of false positives FP represents the number of edges in the NSD map-matched trajectory that do not belong to the corresponding ground-truth GPS map-matched trajectory; (iii) the number of false negatives FN represents the number of edges from the ground-truth GPS map-matched trajectory that do not belong to the NSD map-matched one.

The figures associated to these metrics indicate that the combination of TRANSIT with the map-matching step significantly increases even more the accuracy of the mobility information contained in the NSD data. Specifically, when no prior knowledge about the transport mode was assumed to be available, we obtained an overall matching rate (MR) of 63%, and the geographical error (G_e) was 110 meters on average for the complete set



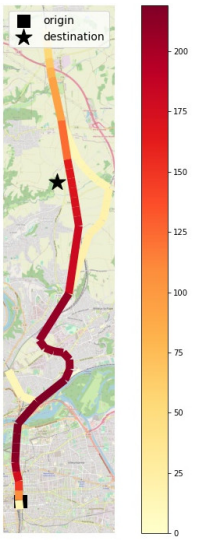
(a) C_1 - Reconstructed popular paths - PT

(b) C_1 - Ground-truth popular paths - PT



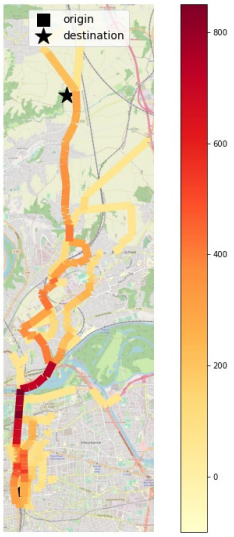
(c) C_1 - Reconstructed popular paths - Road

(d) C_1 - Ground-truth popular paths - Road



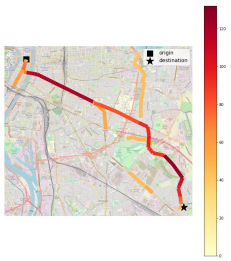
(e) C_2 - Reconstructed popular paths - Train

(f) C_2 - Ground-truth popular path - Train



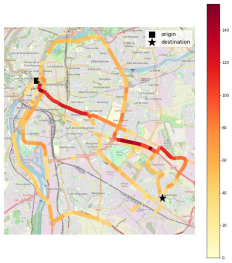
(g) C_2 - Reconstructed popular paths - Road

(h) C_2 - Ground-truth popular paths - Road

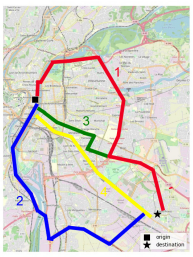


(i) C_3 - Reconstructed popular paths - PT

(j) C_3 - Ground-truth popular path - PT



(k) C_3 - Reconstructed popular paths - Road



(l) C_3 - Ground-truth popular paths - Road

Figure 3.12: Comparison between popular paths reconstructed by our approach and ground-truth popular paths for three case studies: C_1, C_2 and C_2.

of mobile sessions. This level of spatial precision represents a first enhancement compared to the results in Sec. 3.2.3. The results improve significantly when the mode of transport of each trip is provided to select the sub-graph for the map-matching step, with a nearly halved G_e of 60 meters and a 10% improvement in the MR, reaching a satisfactory value of 77%. Furthermore, the map-matching approach achieved superior results in public transport scenarios than on roads, with values as low as 30 meters and as high as 86% for the metrics G_e and MR, respectively, when the prior knowledge of the transport mode is used. The reason for this outcome is the more complex topology of road networks compared to public

transport, making the map-matching problem more challenging in the former case.

We conducted additional validation by processing the entire dataset $\mathcal{D}_{\mathcal{L}}$ from Sec. 1.1.2. We used TRANSIT to extract all (possibly enhanced) trips for a limited subset of OD pairs. For each selected OD pair, we identified the most common paths from the map-matched TRANSIT-reconstructed trajectories, by isolating the transportation network edges with the highest number of occurrences from the total set of map-matched trajectories. We also retrieved typical multi-modal itineraries for the same OD pairs by exploiting a variety of route planners[15], using the latter as ground truth. The results are graphically presented in Fig. 3.12, indicating an overall good match for most of the OD pairs considered.

In conclusion, the analysis demonstrated the potential for obtaining accurate route-level mobility information by transportation mode via network signalling data using simple, coarse transportation mode inference. Our approach can be applied at the scale of an entire country, and with the possibility of frequently recomputing mobility patterns over time or even in real-time. However, the map-matching approach still has many limitations, especially in terms of its limited ability to automatically perform transport mode detection. Several features, such as the probability that the Viterbi algorithm generates as output, the start time/duration of the trip, and more realistic transition probabilities based on time-dependent travel time information on the links of the transportation networks, could be utilised to improve the proposed approach and achieve even higher accuracy.

3.4 Towards Privacy-aware Mobility Inference

TRANSIT provided clear evidence of the necessity for special care in the handling of mobility data retrieved from massive NSD, as they might contain sensitive personal information. We take as a reference the GDPR European regulation, which defines as personal any piece of information pertaining to a particular individual. Based on this definition, mobility trajectories are very sensitive because of their frequent uniqueness. Individuals tend to have very different trajectories over the course of a day, which makes it difficult to extract trends or group them in order to provide more privacy-preserving mobility reports [137]. This problem applies anyway even to sparser and less pervasive sources of data such as CDR, LBSN, GPS and even surveys, which can often provide unique trajectories or indicators of individual mobility.

Let us consider for instance the case of a dynamic OD-matrix, a crucial indicator of mobility describing the flows between origins o and destinations d over a time window $[t, t + \Delta t)$. Typical values for Δt range from 15 to 60 minutes, depending on the specific transport application and the data sources used to produce them [168]. A set of OD-matrices is then a regular, relational dataset with three attributes for each flow: an origin o, a destination d, and a time t usually denoting the time step of the beginning of the flow. Although they represent a dramatic simplification compared to trajectories, they still preserve high levels of uniqueness. Flows between specific pairs of zones can be in fact very small and isolated from others, and this problem is amplified when the adopted spatio-temporal resolution is finer[16]. In such scenario, OD-matrices are characterised by a high number of modalities[17], *i.e.,* the origin and the destination can be selected from sets of up to thousands of areas. These features make OD-matrices thus harder to anonymise than regular relational data.

To address this challenge, we proposed in [169, 170] a methodology to efficiently make dynamic OD-matrices meet GDPR's definition of anonymous data, *i.e.,* the individual is

[15] https://www.google.fr/maps and https://www.viamichelin.fr/web/Itineraires

[16] As discussed in Sec. 3.2.4, TRANSIT can be used to generate such indicators at very detailed spatio-temporal resolutions. We tested TRANSIT to reconstruct OD flows at spatial resolutions lower than 1 km^2 and with temporal binning in the order of minutes.

[17] We call modality each pair of OD zones deriving from the adopted segmentation of the territory under analysis.

not or no longer identifiable (as from Recital 26 of the GDPR). The widely used criterion of k-**anonymity** [171] was adopted to achieve this objective. An OD-matrix is said to be k-anonymous if no flow represents less than k individuals. By k-anonymising an OD-matrix, we can improve its safety against various categories of attacks [172], namely:

- **Record linkage**, also called re-identification attack. If a flow in the OD-matrix has only one individual, an attacker can pinpoint a target individual in the data.
- **Attribute linkage**, also called homogeneity attack. If all the individuals from a particular origin go to the same destination at a given time, then an attacker knowing that a target left the origin area can infer where they went, without needing to pinpoint the target first.
- **Probabilistic attack**. If an attacker knows anything about a target, then accessing the OD-matrix can improve their knowledge about the target's whereabouts.

Each one of these three attacks is a relaxation of the previous one, the goal being less ambitious for the attacker but the success more likely. Note that probabilistic attacks may be successful even if the target user is not actually in the data, as long as we consider the OD-matrix to be representative of the population flows in the study area.

Technically, k-anonymity with k as low as 2 is secure against record linkage attacks, although this would understandably not be a satisfying anonymity condition. In particular, it would offer very limited protection against attribute linkage and probabilistic attacks. State-of-the-art usually aims at k between 3 and 5 in the case of hard-to-anonymise datasets, and up to 200 for simple datasets [173][18].

We used an approach to k-anonymisation based on **generalisation and suppression** [171], *i.e.,* reducing the granularity (*i.e.,* precision) of the data until the flows are big enough (*i.e.,* larger than k) and suppressing the ones that are not. A toy example of 10-anonymisation generalisation and suppression is illustrated in Fig. 3.13: flows $A \to A$, $A \to B$ and $A \to C$ are aggregated together in order to reach a volume above 10, and similar aggregations are performed on the flows originating from C and D, and directed towards destinations B, C and D. Flow $B \to A$ is suppressed, which can be preferable when a very coarse aggregation becomes necessary to hide this flow among others.

Finding such a solution that minimises the loss in precision is known to be a NP-hard problem [176]. Historically, the first k-anonymisation algorithm was Datafly [171], which relies on a generalisation hierarchy describing how modalities should be merged together. An example of such a generalisation hierarchy is illustrated in Fig. 3.14: for a dataset giving the position of individuals in a study zone, the initial modalities are represented by the leaves of the tree. If we choose to generalise the whole dataset one level higher, then the possible modalities are the parents of the leaves. Datafly finds the best horizontal cut in the hierarchy, meaning all individuals (in our case, the OD flows) in the data are generalised to the same level, independently on the area they are localised into. This **uniform generalisation** approach has the advantage of being scalable to huge volumes of data, as well as data with numerous attributes. Within this category, the OIGH algorithm [177] represents the latest and best adapted solution to an OD-matrix. In a bid to find a finer-grained result, some approaches look for a generalisation at the individual level, which gives a solution akin to a clustering [173]. In the specific field of mobility data, this approach is better

[18] A complementary property for data anonymisation is l-diversity [174], which holds when each generalised value (*e.g.,* the aggregate flow encompassing multiple modalities) covers at least l distinct modalities. Together, k-anonymity combined with l-diversity ensures protection against attribute linkage attacks, and significantly reduces the potential of probabilistic attacks. For even further protection, it is possible to implement t-closeness [175], which holds when the distribution of attributes in each group of k is no farther than a threshold t from the distribution of the overall population. However for OD-matrices, it can be argued that k-anonymity alone, given a sufficiently high k, is enough to make the individuals no longer identifiable: the areas of origin and destination can be large enough to avoid attribute linkage, and the low number of attributes renders probabilistic attacks rather vacuous.

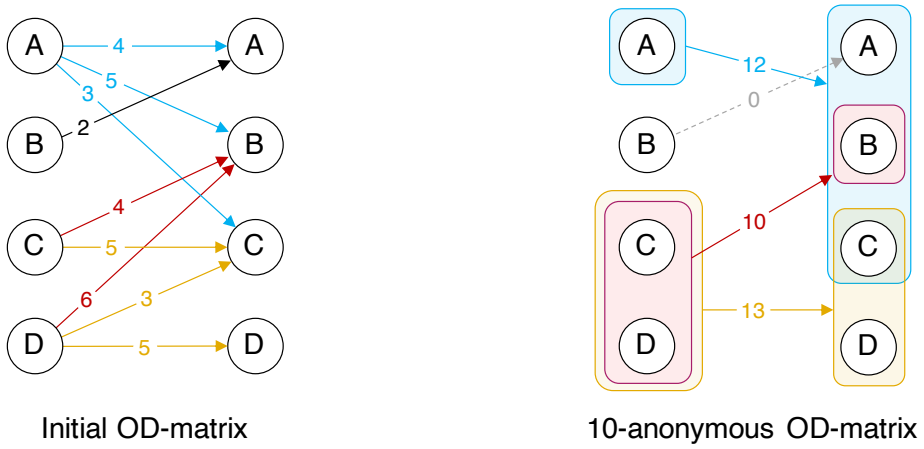


Figure 3.13: Example of generalisation and suppression of a simple OD matrix. Note that the flows here have been clustered without any kind of constraint. Some approaches seek particular solutions, notably ones where the generalisations form a partitioning of the domain.

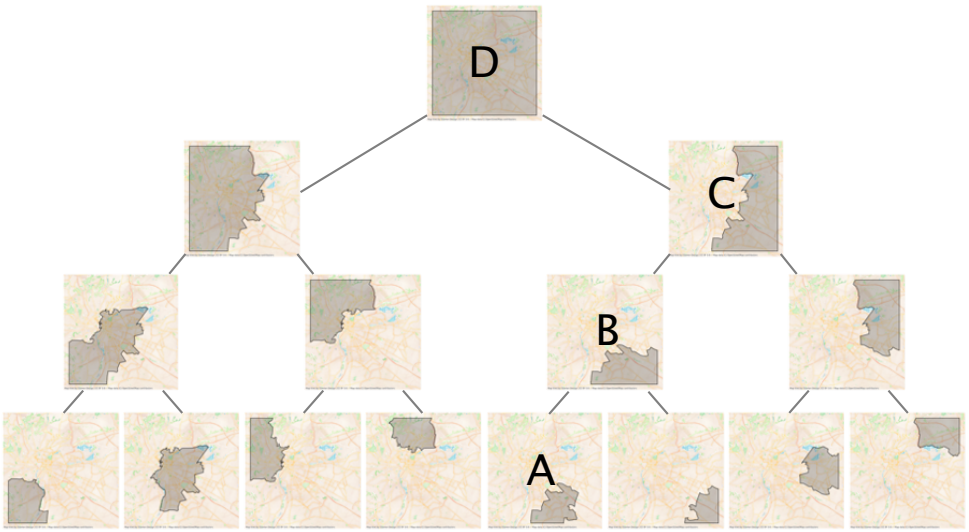


Figure 3.14: Example of a spatial generalisation hierarchy. The root represents the whole study area, and the children of a node form a partitioning of the parents. An individual present in area A in the data can be generalised to be shown as present in area B, C or D depending on what is necessary in order to hide them in a group of k individuals.

represented by Glove [137], which generalises points in a dataset of trajectories. However, these approaches lack scalability for datasets characterised by a huge number of flows, as is the case for data derived from mobile phone CDR and NSD: for example, Liang and Samavi [173] report computing times in the order of hours for low values of k, and their approach can be expected to take days for $k \geq 10$ given its exponential time dependency. Other solutions for general-purpose anonymisation rely on the notion of differential privacy, a robust privacy principle introduced by Dwork et al. [178]. Differential privacy does not apply to a dataset in itself but rather to a randomised algorithm taking the dataset as input and returning a series of query results. An algorithm is said to satisfy differential privacy if the probability of obtaining any output is not significantly different regardless of whether

an individual's data is included or not in the dataset.

To achieve rapid and accurate anonymization of OD-matrices, we developed an adaptive approach, formulated as optimisation problem, that utilises generalisation and suppression over a tree-based hierarchy. The problem aims at finding optimal partitionings (*i.e.,* groupings) of the origin and destination areas that minimise the amount of introduced generalisation, with the constraint of satisfying the generalisation hierarchy and keeping low the number of suppressed records to a fixed maximum threshold. This last constraint is critical to guarantee the representativeness of the produced OD matrices for useful transportation applications. A comprehensive description of the optimisation problem can be found in [170] for interested readers.

We employed state-of-the-art algorithms for dependency-constraint knapsack problems [179, 180] to solve the optimisation problem and developed an adaptive generalisation technique that provides more refined results than uniform generalisation. Using this formulation, we proposed the Adaptive Tree Generalisation (ATG) approach, a lightweight algorithm that efficiently achieves k-anonymity. The ATG algorithm was developed in two versions, namely, ATG-Dual and ATG-Soft, which correspond to two variations of the problem we solved.

We evaluated our approaches against an extensive benchmark of anonymisation methods from the state of the art: general-purpose solutions from the field of anonymisation (*i.e.,* **OIGH** [177] with uniform generalisation and **Mondrian** [181]), clustering from mobility data anonymisation (*i.e.,* **Glove** [137] and our custom **Glove-sk** simplified implementation of Glove for large-scale datasets), and differential privacy (*i.e.,* **Laplace Noise**). We also considered in the benchmark the results obtained by performing **suppression** alone of all flows that do not respect the k-anonimity constraint.

The approaches have been compared on a variety of datasets: New York City taxis (`nyc`), available as open-data[19], and the Senegal (`senegal`) and Cote d'Ivoire (`civ`) datasets, available in the scope of the Data for Development (D4D) challenge [104, 103]. Several variations on the `senegal` dataset were also considered: `senegal_crop` with only the 599 eastern-most base stations, which account for most of the activity; `senegal_big` which increases the volumes in the data by summing together flows related to periods pertaining to different, but possibly overlapping, sets of individuals; and `senegal_split` which artificially increases the number of initial tiles by dividing the base stations in four and distributing the flows non-uniformly between them.

name	#matrices	#tiles	average #flows	density	average volume	%anon. flows	%anon. vol
nyc	7242	263	3058	4.4%	15009	15.8%	59.0%
civ	632	1221	3523	0.2%	3117	0.3%	2.6%
senegal	6752	1666	18276	0.7%	41027	5.7%	36.8%
senegal_crop	6528	599	13710	3.8%	29614	5.8%	37.8%
senegal_split	360	6664	305194	0.7%	956742	6.5%	46.0%
senegal_big	360	1666	100322	3.6%	956742	12.3%	80.7%

Table 3.3: Descriptive statistics of the datasets used for the experiments (#matrices: number of matrices in the whole dataset, where each matrix represents the flows over a time step; #tiles: number of initial tiles over which the matrices are set; density: average graph density of the matrices; avg. #flows: average number of flows among the matrices in the dataset; avg. vol.: average sum of the flows; %anon. flows: fraction of flows that are above $k = 10$; %anon. vol: fraction of individuals that are in flows above $k = 10$.)

Table 3.3 summarises the main characteristics of the datasets: we detail in particular the *density* of the matrix, computed as the number of non-null flows over the total possible number of distinct flows (*i.e.,* the squared number of tiles), averaged over the whole period of observation, and the *percentage of flows that are 10-anonymous* in the original dataset.

[19] publicly available at: https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page

The relatively low amount of these 10-anonymous flows corresponds to a significant share of the total volumes of the data, as they are naturally associated with the OD pairs traversed by the most significant amount of traffic. The part of the anonymous volume in the original data is given under the column *%anon_vol*.

For the choice of performance indicators, we assumed that the generalised (*i.e.,* aggregate) areas provided as output by the anonymisation approach will most likely not be of interest for the final users of the data. Therefore, it appeared reasonable to assume that data users would rather be interested in drawing their own areas of interest and querying the matrix for an estimation of the flow between these areas. In the absence of additional information, the best estimation for such a custom flow is obtained by considering the volumes to be uniformly distributed in their generalised areas, and, as a result, the flow between the arbitrary areas can be derived proportionally to the overlaps with the intersected generalised areas. We call this estimation process **reconstruction**.

It is then relevant to actually *reconstruct* the anonymised OD matrix over the original segmentation and evaluate the difference with respect to the original OD matrix. The error metric E reported in the results is thus the absolute difference across all flows of the reconstructed anonymised matrix and the original one, *i.e.,* formally:

$$E = \frac{1}{V} \sum_{o,d \in leaves(T)} |\tilde{v}_{o \to d} - v_{o \to d}| \tag{3.36}$$

where $\tilde{v}_{o \to d}$ denotes the reconstructed volume over the initial tiles (*i.e.,* the leaves of the generalisation hierarchy), and $V = \sum_{o,d \in leaves(T)} v_{o \to d}$ normalises the error by the total volume of observed flows. Note that the sum is over the leaves of T, corresponding to the initial tiles over which the OD matrix is defined.

We also evaluate the approaches with respect to their total generalisation error $\bar{G}$, defined as follows:

$$\bar{G} = \frac{1}{V^+} \sum_{o \to d \in \mathcal{F}^+} (|o| + |d|) v_{o \to d} \tag{3.37}$$

where $\mathcal{F}^+$ represents the set of anonymous flows $o \to d$ such that $v_{o \to d} \geq k$, and $V^+ = \sum_{o \to d \in \mathcal{F}^+} v_{o \to d}$ the total volume of anonymised flows. When the origins and destinations are aggregated to roughly the same level as is normally the case, $\bar{G}$ represents roughly twice the number of tiles in the origin or destination of the average generalised flow. A value of $\bar{G} = 2$ then means that no generalisation was performed. We also consider the fraction of suppressed volumes S as an indicator of accuracy.

Finally, as a complementary metric to measure the distortion induced by the anonymisation, we evaluate the **distribution distance** D between each OD matrix and its anonymised versions. This metric is slightly different than the reconstruction loss E as it considers OD-matrices as normalised distributions, not penalising suppression. The distribution distance is given by:

$$D = \sum_{o,d \in leaves(T)} \left| \frac{\tilde{v}_{o \to d}}{V^+} - \frac{v_{o \to d}}{V} \right| \tag{3.38}$$

with the same notation as the previous definitions.

Table 3.4 summarises the average performance over all matrices in the small datasets `nyc`, `civ`, `senegal` and `senegal_crop`, for which we were able to run Glove. We report the mean value of $\bar{G}$, E, D and S, and the total computing time to anonymise all the matrices. Firstly, Glove stands out for its lack of scalability, making it impractical in huge-volume cases.

Among the other solutions, the difference of a few minutes in the computing time is relevant but not decisive. They must instead be compared based on $\bar{G}$, for which ATG-Dual

solver	$\bar{G}$	E	D	time (s)	S
ATG-Dual	24.90	1.18	1.14	23	9.99%
ATG-Soft	40.41	1.76	1.31	9	3.24%
oigh	53.62	1.34	1.30	56	9.87%
glove-sk	444.12	1.10	1.36	358	9.99%
glove	18.70	0.81	0.84	23203	9.98%
mondrian	27.39	0.95	1.02	688	0.00%
suppression	—	0.33	1.34	—	39.46%
laplace_noise	—	2.88	6.78	—	-170.92%

Table 3.4: Performance on samples of the dataset senegal_crop, nyc, civ, and senegal. $\bar{G}$: mean generalisation error; E: normalized reconstruction loss; S: fraction of volumes suppressed; D: distribution distance (Eq. 3.38). The reported time is the total computing time (in seconds) to run the anonymisation of all matrices of all datasets. Note that laplace_noise adds volumes, as it mostly applies a positive noise on a sparse matrix.

solver	$\bar{G}$	E	D	time (s)	S
ATG-Dual	**19.85**	**0.72**	**0.78**	111	9.92%
ATG-Soft	65.39	1.07	1.09	**32**	1.54%
oigh	29.68	1.10	1.15	985	9.79%
glove-sk	1361.51	1.04	1.25	3376	10.00%
mondrian	25.19	0.76	0.85	11541	0.00%
suppression	—	0.49	1.02	—	59.00%
laplace_noise	—	2.87	3.33	—	-86.98%

Table 3.5: Performance on samples of the datasets senegal_big and senegal_split.

significantly improves state of the art, with Mondrian representing the second best solution, with a very good E and D – primarily due to the fact that, differently from ATG approaches, it does not suppress any volume – and a reasonable computation time.

For differential privacy, E and S mostly measure the volumes that have been added to 0-flows, as they represent more than 95% of possible flows in each OD-matrix. Because of this, differential privacy performs worse than any generalisation technique when compared based on the E metric. This matches the previous observation [182] that differential privacy is not adapted for sparse data. As the interpretation of $\bar{G}$ is only relevant for generalisation, we do not include it in the table for suppression and laplace_noise that do not use this technique.

As Glove could not be evaluated for our biggest datasets, we compared the approaches in a separate Table 3.5 for `senegal_split` and `senegal_big`. We see that the best computing time we could hope for Glove, given by Glove-sk, is still not satisfactory in this situation. Our approach is of particular interest here, as it offers a finer generalisation than OIGH and Mondrian for a fraction of the time. The ATG-Soft alternative is even faster, making it relevant for anonymising even bigger matrices, such as the total amount of base stations of a highly urbanised country.

Even if our approach is more appropriate, uniform tree generalisation admittedly performs well for OD matrix generalisation. Indeed, we could expect the best solution to show a high disparity of aggregation levels between densely and sparsely populated areas, which a uniform generalisation cannot offer. This effect is mitigated by the fact that the initial tiles already partially reflect the disparity in activity density, as they rely on base stations or administrative divisions, which are more densely distributed in populated areas. This illustrates the importance of the generalisation hierarchies, but it is also partially explained by some code adaptations we implemented for OIGH to run on hierarchies whose initial leaves are not all at the same depth.

In conclusion, the proposed methods make a valuable contribution to efficiently anonymising large trajectory datasets in the form of dynamic OD matrices, while preserving input

data resolution. Our approaches facilitate the development of aggregate mobility indicators that enable fine-grained studies on human mobility while maintaining fundamental privacy requirements. However, it is important to note that k-anonymity alone can be insufficient to protect against attribute linkage and probabilistic attacks. To achieve complete privacy security, the method should also guarantee l-diversity and t-closeness. Nevertheless, the k-anonymity provided by our approach is a necessary initial step towards a more comprehensive solution and meets the current requirements of the French regulator CNIL, responsible for enforcing GDPR in France. Presently, mobility data are underutilised despite their enormous potential, largely discussed in this chapter, due to their size and personal nature, which makes their handling expensive and legally risky. Implementing an affordable, quick, and reliable anonymisation process should thus go in the direction of facilitating their widespread public use and promoting the extraction of useful insights to optimise and make more resilient transportation networks, as detailed in the next chapter.

Chapter 4

Assessing and Managing Vulnerabilities in Transportation Networks

Chapters 2 and 3 reported our findings on how massive data can provide accurate insights into the dynamic performance and actual demand of a large-scale, multi-modal transport system in both recurrent and non-recurrent situations. These insights contribute to *learning from the past*, a crucial aspect of resilience as discussed in Sec. 1.2. Specifically, we demonstrated the efficacy of using passively-collected mobile phone data, particularly NSD, to reliably reconstruct both aggregate and fine-grained spatio-temporal estimations of travel demand and traffic variables in the form of OD matrices, regional speed dynamics, travel mode choices, popular paths, *etc.*

Although having a proper understanding of travel demand and traffic performance based on past observations is crucial for planning robust and efficient transportation systems, such knowledge alone is insufficient for effectively protecting these systems from disruptive events, for which data may be scarce or completely unavailable. Such events can lead to severe declines in the quality of mobility services and impact vulnerable segments of the transportation system and other urban infrastructures, thus potentially causing cascading failures and significant economic and social losses.

To reduce the consequences of disturbances and infrequent events, *vulnerability assessment* enables the analysis of how engineered systems, particularly cyber-physical ones, respond to specific kinds of disruptions. As detailed in Sec. 1.2.2, vulnerability assessment corresponds to the *anticipating* (*knowing what to expect*) cornerstone of resilience. It aims to identify risk situations before they occur and spot the system's components that might be unable to cope with such events, thus requiring special attention and enhancements that can increase the system's robustness and recovery ability. State-of-the-art studies on transport vulnerability traditionally neglect the complexity of multi-modal transport topologies and the impact that traffic conditions and travel demand dynamics can have on the system's response to a disruptive event [66, 38].

Additionally, extreme weather and human-related events represent a fundamental risk category that increasingly challenges transport systems and whose impact on such networks can be hard to model and anticipate. Nevertheless, with proper metrics and real-time data supporting continuous monitoring of the entire transport network, its surroundings and its most critical components, detecting such events rapidly and deploying appropriate strategies to mitigate their consequences may become possible. These aspects are related to enhancing the *monitoring* (*knowing what to look for*) cornerstone of transport resilience.

We also denoted a scarcity of operational decision-making tools to support resilient-driven

planning and management of multi-modal transport networks by considering the system's exposure to both recurrent and unexpected disruptive events. These tools correspond to the *knowing what to do* cornerstone of resilience and are essential for properly *responding* to frequent disruptions and unlikely events, thereby enabling transport networks to maintain adequate operations even during abnormal situations.

Based on these observations, this chapter summarises the main contribution of my research on transport vulnerability assessment, monitoring and management, aimed at answering the following questions: *(i)* What are the most appropriate metrics to identify vulnerabilities of a transport network? *(ii)* What approaches can be leveraged to model the occurrence of disruptions, evaluate the response of the transport network, and measure the consequences of such disruptions in terms of performance? *(iii)* How can data-driven, fine-grained, and time-varying knowledge of travel demand and traffic-related variables be integrated into the identified approaches for more accurate and rapid detection of disruptions and network vulnerabilities on a large scale? *(iv)* How can new on-demand mobility services be optimally designed and integrated into an existing transport network to reduce vulnerability to disruptive events?

The contribution to these research questions is detailed in this chapter as follows. Sec. 4.1 outlines the findings of a comparative study of approaches and metrics for vulnerability assessment. This study demonstrates the main limitations of traditional system-based and graph-based topological approaches, highlighting the need for novel combined strategies that can enact real-time monitoring of transport networks. Based on these findings, Sec. 4.2 proposes an original approach combining big data processing and complex network metrics for promptly detecting and possibly anticipating disruption occurrence, thus being helpful in implementing real-time monitoring of complex and dynamic networks. This study laid the ground for the research on real-time monitoring of large-scale transport networks, which will be detailed in Chapter 5. Finally, Sec. 4.3 discusses our solution based on multi-modality and stochastic optimisation to support the resilience-oriented design of transport networks.

The reported research activities have been performed in the context of the ANR JCJC PROMENADE project and have been at the core of Elise Henry's PhD thesis on resilience modelling of urban multi-modal transport networks, which I co-directed.

This chapter includes content from the following papers:

1. P. Gauthier, A. Furno, and N.-E. El Faouzi, "Road network resilience: how to identify critical links subject to day-to-day disruptions", Transportation Research Record, vol. 2672, no. 1, pp. 54–65, 2018.

2. E. Henry, A. Furno, and N.-E. E. Faouzi, "Approach to quantify the impact of disruptions on traffic conditions using dynamic weighted resilience metrics of transport networks", Transportation Research Record, vol. 2675, no. 4, pp. 61–78, 2021.

3. E. Henry, L. Bonnetain, A. Furno, N.-E. El Faouzi, and E. Zimeo, "Spatio-temporal correlations of betweenness centrality and traffic metrics", in 2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS). IEEE, 2019, pp. 1–10.

4. E. Henry, A. Furno, N.-E. El Faouzi, and D. Rey "Locating park-and-ride facilities for resilient on-demand urban mobility", Transportation Research Part E: Logistics and Transportation Review, vol. 158, 102557, 2022

The research reported in this chapter is strongly related to the activities discussed in Chapter 5. In particular, Chapter 5 describes the development of the PROMENADE architectural and technological platform, which leverages cloud, IoT, micro-services and big data solutions to provide the necessary support for enhancing resilience through real-time monitoring and control in intelligent large-scale urban environments.

4.1 Vulnerability Analysis: a Comparative Study

In Sec.1.2, we introduced the limitations of topological and system-based approaches to the vulnerability assessment of transport networks. These limitations were empirically confirmed in our work[55], in which we proposed a novel system-based methodology, termed the *stress-test* approach. This approach is based on dynamic mesoscopic simulation and aims to identify the most critical links within a road network. The methodology was conceived to model both day-to-day recurrent disruptions, such as accidents and mild weather-induced traffic degradation affecting road link capacity, and extreme events leading to sudden and significant variations in travel demand. The results obtained using this methodology were compared with those garnered using topology-based static metrics. This comparison highlighted the limitations of the latter from a functional perspective and those of the former from a performance perspective.

From a general point of view, stress testing consists of evaluating a system's performance under extreme conditions that go beyond normal usage, usually by relying on simulation. Stress testing helps identify a system's weaknesses or failure points and determine how it can be made more resilient. It has been widely used in finance [183, 184], as well as in the medical [185, 186] and the hydro-geology domains [187]. For instance, banks might conduct stress tests to evaluate the resilience of lending portfolios by modelling the impact of severe economic junctures on loan defaults. In the field of transportation, stress testing has been traditionally performed via approaches that remove edges or nodes from a network to model the occurrence of a (spatial) disruption, using performance indicators, *e.g.,* based on total travel time increase, to evaluate the impact of the modelled disruption [46, 188, 189, 82].

The large majority of research works on stress testing of road networks focuses on major disruptive events [189, 190, 191], ignoring day-to-day disruptions that negatively and very frequently affect the performance of traffic networks. It is also worth noting the lack of studies combining system-based with topological approaches, as already mentioned in Sec. 1.2.2. Even though weighted graph-based models and metrics that include traffic flow data exist in the literature, their applicability to system-wide, dynamic traffic models is still neither widespread nor fully understood, being these metrics often inaccurate in highly dynamic environments as well as prohibitive to compute on large scale networks [61, 192].

The research described in this section was motivated by these considerations and represented our first step in the direction of joining solutions from the two categories of vulnerability assessment approaches (*i.e.,* topological and system-based stress-testing), as further detailed in Sec. 4.2.

4.1.1 Graph-theory metrics

The topological metrics used in this study were based on Betweenness Centrality (BC), originally proposed in [193]. BC measures how central a link in a graph is by considering the number of the shortest paths that pass through the link, relative to the total number of shortest paths connecting any given pair of nodes. As highlighted in Sec. 1.2.2, BC represents one of the most widely used metrics in the literature to perform transport network vulnerability analysis. Edge BC is formally defined as follows for the generic link $l \in E$ of a graph $G(V, E)$:

$$BC(l) = \sum_{i \neq l \neq j} \frac{\sigma_{ij}(l)}{\sigma_{ij}}, \tag{4.1}$$

where:

- $\sigma_{ij}(l)$ is the number of shortest paths from node i to node j that traverse link l;
- σ_{ij} is the total number of shortest paths from node i to node j.

An equivalent formula, replacing the generic link l with the generic node n, corresponds to the formal definition of *node* betweenness centrality, $BC(n)$. In shortest path computation, links can be unweighted or weighted, as for example by the associated estimated travel cost (*e.g.,* travel time). In this study, we tested and compared both cases.

In order to model different aspects of a road traffic network and accelerate the computation of BC on larger networks, we also considered multiple variants of the BC, mostly consisting in calculating the shortest paths from *travel demand entry* to *travel demand exit* nodes only[1]. This definition introduces two advantages: computation time is reduced; the metric becomes more realistic from a demand-aware perspective since individuals tend to start and finish their trips over a subset of road intersections. The formula is the same as Eq. 4.1 with the following exceptions:

- i is selected from the entry-nodes subset, *i.e.,* a limited number of intersections where vehicles enter the road network;
- j is selected from the exit-nodes subset, *i.e.,* a limited number of intersections where vehicles leave the road network.

In conclusion, we considered four different formulations of the edge BC:

- Unweighted Betweenness Centrality (UBC)
- Travel-Time weighted Betweenness Centrality (TTBC)
- UBC on entry/exit nodes only (UBC entries-exits)
- TTBC from entry to exit nodes only (TTBC entries-exits)

It is worth remarking that we used static travel time information, *i.e.,* travel time in free flow conditions, to compute both TTBC and TTBC entries-exits.

4.1.2 Demand-sensitive Vulnerability Metric

Jenelius *et al.* introduced in [78] the demand-aware metric of link *Importance* (I) to characterise transportation vulnerability. This metric allows measuring the network performance loss induced by a disrupted link l, by using a generic metric of travel cost weighted by the traffic demand. Differently from [78] that uses a binary approach in modelling link disruptions (*i.e.,* a link is either fully operational or completely closed), we considered the possibility that the capacity of a link (*i.e.,* the maximum number of vehicles that can traverse the link) might be partially reduced of a percentage δ of the original capacity, due to accidents or lane closure. This technique allows considering different possible intensities of the disruptive event on a given link, which is especially helpful to model day-to-day disruptions. Such a metric appeared adequate in the scope of our methodological approach as it captures travel demand and the dynamic phenomenon of congestion (*i.e.,* travel costs increase when the network is congested).

In our study, we considered travel cost on a given origin-destination pair as the travel time to move from the origin to the destination divided by the associated travel distance. It is therefore measured in seconds/kilometres. The importance of a link l is thus defined in our work by the following equation:

$$I(l,\delta) = \frac{\sum_i \sum_{j \neq i} x_{ij}(c_{ij}^{\delta}(l) - c_{ij}^{0})}{\sum_i \sum_{j \neq i} x_{ij}}, \tag{4.2}$$

[1] In a transportation network, not all nodes are responsible for generating and absorbing travel demand. Entry and exit points serve as specialised nodes that naturally associate with some level of travel demand. Their identification is generally tied to the definition of OD matrix zones and their urban topology, pinpointing specific nodes within each zone where trips can either commence or conclude.

where:

- x_{ij} is the demand from origin node i to destination node j (measured as number of vehicles);
- $c_{ij}^{\delta}(l)$ is the mean travel cost (measured in seconds/kilometres) from origin node i to destination node j when link l is disrupted at level δ;
- c_{ij}^{0} is the mean travel cost (measured in seconds/kilometres) from node i to node j in the base case, *i.e.,* no disruption present in the network.

To model traffic dynamics and calculate travel costs on a vehicle basis, we used a dynamic mesoscopic simulator from our research team [194], based on the Lighthill-Whitham-Richards model [195, 196] and a dynamic traffic assignment procedure for distributing vehicles along the alternative paths in the network, according to the traffic conditions at the moment the vehicle is generated.

4.1.3 Stress Testing Methodology

Our methodology for road network stress testing aims to identify the most critical links in the road network that are considered vulnerable to different kinds of disruptions. The methodology is composed of the following steps:

1. *Simulating disruptive road events*:
   We proposed two strategies to perform this step. In the first, similarly to [82], we simulated day-to-day road disruptions as link capacity drops. The *capacity-disruption level*, denoted as δ in the following, is defined as the reduction in link capacity, expressed as a fraction of the original one. As anticipated, we considered the possibility of modelling partial capacity drops and analysing the evolution of the road network performance depending on the capacity-disruption level. For simplicity, we only allowed 5 possible capacity-disruption levels on each examined link, *i.e.,* $\delta \in \{0\%, 20\%, 40\%, 60\%, 80\%\}$. Ultimately, the following equation allows computing the maximum capacity of the generic link l in the presence of a capacity-disruption level δ:

$$\delta = 100 \cdot \left(1 - \frac{q_{max}^{\delta}(l)}{q_{max}^{0}(l)}\right), \tag{4.3}$$

   where:

   - δ is the capacity-disruption level applied to link l (percentage) with $\delta \in \{0\%, 20\%, 40\%, 60\%, 80\%\}$;
   - $q_{max}^{\delta}(l)$ is the capacity of link l (measured in vehicles/hours) when it is disrupted at level δ;
   - $q_{max}^{0}(l)$ is the capacity of link l in the base case (measured in vehicles/hours);

   As a second strategy to simulate disruptive road events, we considered increases in the travel demand on specific entry/exit pairs of nodes of the network. By this approach, it was possible to simulate exceptional situations like city evacuations following extreme events (*e.g.,* flooding, attacks, etc.) that typically put significant strain on the road infrastructure and result in total congestion of the network. This strategy consists in changing the OD matrix, *i.e.,* increasing the traffic flow from given entries to specific exit points, and comparing the stress testing results with the modified demand level.

   Based on the selected strategy for disruptive road events, we set the parameters of our mesoscopic simulator (*e.g.,* link capacity, traffic demand) and we simulated the network in the specific setting. For both strategies, travel costs were collected for all vehicles in order to compute the performance metrics described in the following point.

2. *Computing overall performance loss*:
   This step aims to quantify the consequences of the simulated disruptive road event on the operation of the whole road network. To this purpose, we used the notion of road network performance, measured via the Importance metric from Sec. 4.1.2. The *overall Performance Loss* (PL), when link l is disrupted, was defined as an adaption of Eq. 4.2 to the situation where costs are known per vehicle:

$$PL(l,\delta) = \sum_{v=1}^{n} \frac{c_v^{\delta}(l) - c_v^0}{n}, \tag{4.4}$$

   where:

   - $PL(l,\delta)$ is the overall performance loss when link l is disrupted at level δ (seconds/kilometres);
   - $c_v^{\delta}(l)$ is the travel cost of vehicle v when link l is disrupted at level δ (seconds/kilometres);
   - c_v^0 is the travel cost of vehicle v in the base case (seconds/kilometres);
   - n is the number of vehicles in the network.

   In order to compare links and identify the most critical ones, a unique value of criticality must be associated with each link. To this purpose, we defined a global metric, namely the Stress Test Criticality (STC), to aggregate the multiple loss values derived from Eq. 4.4 for each possible capacity-disruption level (*i.e.,* $\delta \in \{0\%, 20\%, 40\%, 60\%, 80\%\}$) applied to the given link l. The metric definition for the generic link l follows:

$$STC(l) = \int_{\delta} PL(l,\delta), \tag{4.5}$$

   where:

   - $STC(l)$ is the Stress Test Criticality when link l is stress-tested (seconds/kilometres)
   - δ is the capacity-disruption level of link l (percentage)
   - $PL(l,\delta)$ is the overall performance loss (seconds/kilometres) derived from one simulation.

3. *Analysing the results*:
   At the end of our simulations, we obtained the values of the STC metric for all or a subset of the network links. We then computed the graph-based static indicators described in Sec. 4.1.1. This step involves analysing and comparing the rankings of network links produced by both the graph-based metrics and the overall performance loss indicator STC. Links with higher values of these metrics were given higher rankings.

4.1.4 Evaluation

In this section, we present the results of our stress-testing methodology and discuss the link ranking derived from the different selected metrics. We show that the ranking of critical links on the same network can significantly vary when different indicators are used, thus proving that simple modifications of one centrality indicator can have a relevant impact on the capacity of the metric to capture different facets of network vulnerability. Moreover, we discuss the advantages and drawbacks of each different approach in assessing road network vulnerability, towards the definition of new enhanced metrics and approaches for vulnerability identification in large-scale networks.

Our analysis has been performed in two different case studies: a first one related to a simple virtual network, used as a basic test-bed for our approach; and a second one

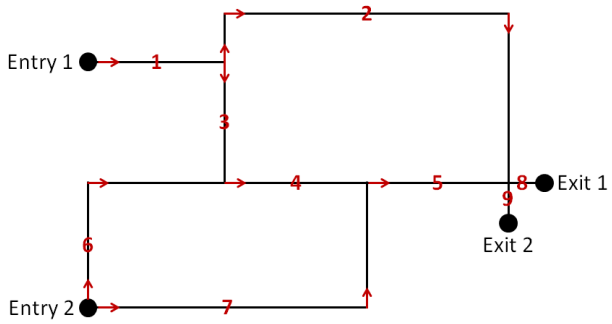


Figure 4.1: Simple virtual road network.

corresponding to a real road network in France, which we used to confirm the validity of our results in a realistic scenario and to support the discussion on advantages and drawbacks of both simulation-based stress testing and topological metrics.

Toy Network Case Study

The simple toy network is composed of 9 links (roads) and 8 nodes. Nodes correspond to 4 road intersections, 2 entry points and 2 exits. The duration of each simulation was fixed at 10 minutes. Fig. 4.1 depicts the network with numbered links and flow directions.

Origin \ Destination	Exit 1	Exit 2
Entry 1	375	375
Entry 2	250	1000

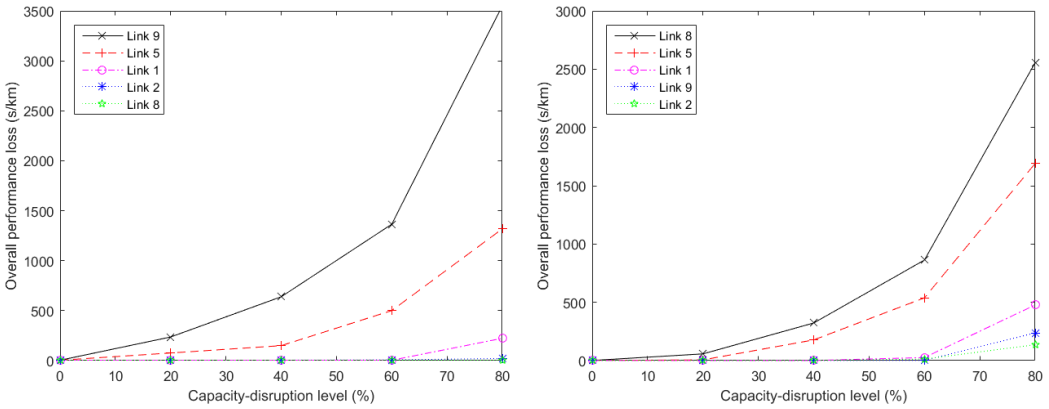
Origin \ Destination	Exit 1	Exit 2
Entry 1	500	500
Entry 2	1200	300

Table 4.1: Origin-destination matrices for the simple virtual network with different demand levels (A and B). Values are expressed as vehicles per hour.

We were able to measure the stress test criticality and all of the proposed topological metrics on all the links of the simple network, due to its limited size. To perform stress testing, we used both strategies for modelling disruptions described in Sec. 4.1.3, *i.e.,* link capacity drop and traffic demand increase. Therefore, two different demand levels, reported as two different origin-destination matrices (A and B) in Tab. 4.1, were defined. The measures of stress-test criticality corresponding to the two different OD matrices are distinguished as STC A and STC B, respectively. It is worth noting that STC A and B are calculated with the same formula (Eq. 4.5) and the same simulation parameters, with the only exception of the travel demand. In our simulations, we applied sequentially five capacity-disruption levels (*i.e.,* 0%, 20%, 40%, 60%, 80%) to each link. Then, we measured the network-wide performance loss (*i.e.,* Eq. 4.4) consequent to the disruptions applied to the link.

The overall performance loss when using travel demand A is reported on the y-axis of Fig. 4.2(a), while the corresponding capacity-disruption levels (δ) are reported on the x-axis. Results for different links are depicted with different colours and markers in the figure, using linear interpolation. For the sake of readability, the figure only reports the five most critical links (*i.e.,* those with the highest overall performance loss). Intuitively, the overall performance loss grows as the capacity-disruption level increases. In other words, a link capacity drop translates into an increase in network-wide travel cost.

According to the definition provided in Eq. 4.5, the stress test criticality of a given link corresponds to the area below the overall performance loss curve. Thus, in scenario A, *link 9* is the most critical to the operation of the whole road network, followed by *links 5, 1, 2* and *8*. By using link ranking from stress test criticality as a baseline, it is possible to compare the other link rankings as derived from the different topological metrics. Tab. 4.2



(a) Scenario A (b) Scenario B

Figure 4.2: Stress testing on a simple test virtual network in two scenarios: performance loss for the top-5 most critical links with different capacity-disruption levels

reports on such link rankings for both stress test criticality and the whole set of topological metrics.

As a preliminary consideration, it can be observed that the rankings of critical links on the same network may dramatically change depending on the selected metric, due to the different properties of the network captured by each of them. As an example, *link 5* is on top of all the topological rankings whereas, in terms of stress-test criticality, it is ranked second, below *link 9*. The top rank of *link 5* with all the topological metrics can be motivated considering the large number of shortest paths traversing this link: *e.g.,* paths *(4, 5)*, *(3, 4, 5)*, *(1, 3, 4, 5)*, *(6, 4, 5)*, *(7, 5, 8)* are all shortest paths.

The different rankings issued by STC A can be easily explained. If *link 5* is disrupted, the alternative paths *(1, 2, 8)* and *(1, 2, 9)* exist for all individuals departing from *entry 1*. Conversely, when *link 9* is disrupted, no alternative path exists for users willing to travel to *exit 2* from both *entry 1* and *entry 2*, thus resulting in significant congestion and consequent travel time increase for all individuals heading to *exit 2*. Additionally, traffic demand for *exit 2* is very high (see OD matrix in Tab. 4.1). That explains why *link 5* is more critical than *link 9* in terms of topology, but less critical than *link 9* when considering demand data, as made possible by our stress-testing methodology (based on dynamic simulations) and captured by the related criticality metric. This simple test clarifies how traditional demand-agnostic approaches may fail in properly ranking edge criticality. It also shows that alternative paths may become the shortest paths of the network when links are disrupted by adverse events, thus attracting traffic flow previously directed through the disrupted links. This is a critical factor that static graph-based approaches fail to capture. However, it is essential to clarify that this does not mean topological metrics are poor indicators of vulnerability. Instead, graph-based models of transportation networks should incorporate dynamic elements derived from real-time data (*e.g.,* traffic sensors, cellular networks, IoT devices, *etc.*) and captured, *e.g.,* via edge weights. Furthermore, metrics like betweenness centrality should be recalculated almost continuously to quickly identify significant disruptions in the network. These aspects are treated extensively in Sec. 4.2.

Another striking difference worth analysing regards *link 2*: it is considered one of the most critical ones according to the UBC metric from entries to exits, whereas it is the least critical one for the TTBC and the TTBC from entries to exits. The peculiarity of *link 2* is its length: it is the longest in the network, thus demanding more time to travel than the other links. Metrics like UBC and UBC from entries to exits are not weighted, *i.e.,* all links are valued equally, and are consequently unable to grasp this important aspect. Differently,

Link	STC A	STC B	UBC	UBC entries-exits	TTBC	TTBC entries-exits
9	1st	4th	2nd	1st	4th	3rd
5	2nd	2nd	1st	1st	1st	1st
1	3rd	3rd	2nd	1st	4th	3rd
2	4th	5th	2nd	1st	7th	5th
8	5th	1st	2nd	1st	4th	3rd
6	5th	6th	4th	2th	5th	4th
4	5th	7th	2nd	2th	2nd	2nd
7	5th	8th	3th	1st	6th	4th
3	5th	8th	3th	2th	3rd	3rd

Table 4.2: Simple network link rankings generated by the different metrics of criticality

links with high travel times are not considered critical by the analysed weighted approaches, because they are not often part of the shortest paths. The same consideration applies to *link 7*, which is the second-longest link of the network.

Finally, it is worth noting that UBC values are often clustered. In particular, the UBC from entries to exits has many equal values and only two different link ranks. Traditional topological metrics appear to have minimal capability to discriminate link criticality at a fine level. In this case, stress test criticality does not differentiate all links either, but this is due to capacity-disruption levels. For *links 8, 6, 4, 7* and *3*, which are all ranked at the 5th place with the same value, demand levels are not high enough, relatively to the other links, to observe a significant performance loss compared to the base case. For example, when capacity disruption is applied to *link 8*, the overall network performance does not change at all (see Fig. 4.2(a)).

To further investigate this aspect, we stress-tested the same simple network with the second travel demand OD matrix, reported as B in Tab. 4.1. Results are shown in Fig. 4.2(b) and Tab. 4.2. Link ranking changes significantly when different demand levels are used. As an example, *Link 8* becomes the most critical link, whereas in the previous case, it involved no performance loss. This is due to the large increase in demand level associated with *exit 1*, which is directly connected to *link 8* (see Fig. 4.1).

Takeaways: *Critical link ranking is highly variable as different approaches are used. Vulnerability assessment via topological metrics is limited in the sense that such metrics do not usually take into account traffic demand and network re-configurations following disruptive events. Conversely, the simulation-based stress testing approach is able to capture these aspects thus providing more realistic rankings via the proposed performance loss metric. Stress testing can also be used to compare different road networks and sub-networks, by analysing their response to similar stresses. TTBC produces better estimations of link criticality with respect to UBC, which treats all links equally.*

Real Road Network Case Study

To confirm the results of our previous analysis in a realistic scenario, we considered a real-world network, the DIRIF, situated in the South of Paris, France. It has 868 links and 827 nodes (657 intersections, 86 entries and 84 exits). Its roads are mostly highways. Each simulation duration was fixed at 15 minutes. As the network is much larger than the toy one, and as traffic flow can be very low on some links, we specified a longer simulation duration to ensure that enough vehicles could travel through the whole network and that we had a proper number of travel cost observations. The simulation was performed with realistic demand data from 9:00 AM to 9:15 AM, corresponding to the morning peak time, in order to increase the probability of observing performance loss as a consequence of the higher demand. The network is graphically presented in Fig. 4.3.

Given the large size of this network and the high computation time associated with

(a) DIRIF, geographic location

(b) DIRIF network

Figure 4.3: The DIRIF road network in Paris agglomeration

	Link	STC	UBC	UBC entries-exits	TTBC	TTBC entries-exits
Highest-flow links	**95**	117	2628	83	2628	83
	93	102	660	83	660	83
	94	101	1974	83	1974	83
Highest-BC links	**802**	42	192497	3029	192717	3039
	803	27	192521	3029	192741	3039
	607	27	192509	3029	192729	3039
	608	44.6	192449	3029	192669	3039
	397	15.8	83164	1139	83164	1139
	672	14.5	10	1	10	1

Table 4.3: DIRIF network link values generated by the different metrics of criticality

each network simulation[2], it was prohibitive to perform an exhaustive stress-test analysis as in the simple network case. Therefore, we decided to perform stress tests on a limited set of representative links: the three links with the highest flow of vehicles in the base simulation with no disruption, the three ones with the highest UBC and three randomly selected edges with UBC in three classes of values (*i.e.,* high, medium and low). We discuss in the following our simulations related to the realistic available values of travel demand for the DIRIF network. Tab. 4.3 reports the actual values of the considered metrics for the analysed links[3].

Consistently with our previous analysis on the simple network, Tab. 4.3 shows that rankings of critical links vary significantly when different metrics are used. As an example, *links 95, 93, 94* have a very high value of STC, whereas the topological metrics rate them much less critical than *links 802, 803, 607*. As pointed out in the previous section, the indications from the STC metric appear to be more reliable, since the metric seems to capture the higher criticality of *links 95, 93, 94* associated with higher demand.

On the small link subset considered in our analysis, taking into account travel times (TTBC and TTBC entries-exits) does not significantly change rankings with respect to

[2]Stress testing one link from such a large network with 5 capacity drops takes more than 1 hour on an Intel Xeon E3 CPU equipped with 8 GB of RAM running a Matlab implementation of the mesoscopic simulator.

[3]Differently from Tab. 4.2, we do not report metric rankings but actual values of the metrics for each analysed link. This was due to the impossibility of performing all the required simulations for computing the STC metric for all network links.

UBC, since lengths (and therefore the free-flow travel times) happen to be very similar for all considered links. Finally, it is worth noting that in the DIRIF network, UBC values (especially in the entries/exits variations) tend to significantly cluster themselves (*i.e.,* many edges have very similar values of UBC), thus exhibiting a lower discriminant power than in the toy-network case study.

Takeaways: *In a real-world scenario, stress testing proved to be a realistic and reliable approach to evaluate network resilience. Our evaluation confirms the importance of traffic demand and network dynamics for fine-grained ranking of the most vulnerable road-network links. Stress testing has however the important drawback of requiring very high execution times due to computationally intensive network simulations.*

4.2 Graph Theory and Big Data Processing towards Online Transport Monitoring

As opposed to purely topological metrics, the system-based stress-testing approach described in Sec. 4.1 confirmed its ability to take into account travel demand and traffic dynamics to reliably identify links that are most critical to day-to-day disruptions and changes of travel demand. In our follow-up work [197], we also extended this approach for handling area-covering disruptive events, similarly to [83].

However, the stress-testing approach has a major drawback: it requires an extremely high number of computationally intensive traffic simulations. Therefore, it is recommended only for vulnerability analysis of small-sized networks, and as a preliminary step in application scenarios that allow for a sufficiently long response time, such as vulnerability-aware network design and maintenance planning (see Sec. 4.3). Another limitation of the stress-testing approach is its limited capability to model complex and hardly predictable disruptions that cannot be solely represented in terms of capacity reduction on single links or areas and/or an increase in travel demand. This is especially true for cascading failures, which combine, usually in sequence, multiple heterogeneous spatio-temporal disturbances causing major anomalies in the mobility system.

In scenarios that require rapid response times, such as recovery of critical infrastructure or emergency evacuations, swiftly detecting complex and unpredictable disturbances through real-time network monitoring is among the few viable strategies. However, effective transport monitoring hinges on access to up-to-date and detailed data on travellers' activity and transportation supply within the network. It also depends on selecting highly responsive metrics capable of quickly identifying or even anticipating the occurrence of disruptive events.

This section describes our research efforts aimed at exploring how topological indicators, specifically, indicators based on degree centrality (Sec. 4.2.1) and betweenness centrality (Sec. 4.2.2), could become suitable metrics for online network monitoring when calculated on a time-varying weighted graph. By this approach, the power of topological analysis combines with the availability of data on a system's dynamics, such as travel demand, speed and traffic flow, the latter being typically associated with system-based approaches. The outcomes of this research have confirmed that continuous data-driven monitoring of specific weighted graph metrics represents a key solution to increasing transport resilience. These studies represent the motivation at the basis of the platform and online tools for real-time monitoring of CIs, which will be described in Chapter 5.

4.2.1 Weighted Degree Centrality Metrics for Transport Network Resilience

In [198], we proposed a dynamic modelling approach combining complex networks and big data processing suitable for online continuous monitoring of road traffic networks.

Specifically, we focused on a specific metric, *i.e.,* network *heterogeneity*, also known as *graph irregularity*, traditionally used as a dynamic indicator of resilience for gene regulatory, ecological and power supply systems. Heterogeneity provides global information on the diversity of network nodes connectivity [199]. When calculated over a time-varying, weighted graph, the heterogeneity metric can account for traffic dynamics, making it sensitive to perturbations. Furthermore, the repeated observation of specific patterns associated with this metric in the presence of disruptions suggests that it could be valuable for the development of effective, real-time monitoring tools. These tools could enable the automatic and rapid detection of such patterns and, consequently, of the associated disturbances.

The definition of heterogeneity is typically based on degree centrality. The degree centrality of a node describes the number of edges (or nodes) that are adjacent to the considered node. On directed graphs, where each edge has a specific direction, it is possible to distinguish between in-degree and out-degree centrality by considering the number of edges entering (respectively, exiting) the analysed node. The definition of this metric can also be extended to the case of a weighted network (*i.e.,* weighted degree centrality), considering the sum of all the edge weights connected to the generic node i, *i.e.,* $C_i = \sum_{j \in N(i)} w_{i,j}$, where C_i is the non-normalised weighted degree centrality of node i, $N(i)$ is the set of neighbours of node i, and $w_{i,j}$ represents the weight of the edge connecting nodes i and j. As with the traditional degree centrality, the weighted degree centrality of the generic node i can be distinguished into in-degree and out-degree centrality, being equal to the sum of the edge weights that join the node i to its predecessors (respectively, successors) j.

Different definitions of network heterogeneity have been proposed in complex network theory. For instance, in [200], the authors use the entropy of nodes degrees, concluding that such metric captures the disorder of the analysed network and represents an effective measure of network resilience to random failures. Indeed, variations of this statistical measure can describe losses of information in a network and thus indicate the impact of disruptions in terms of reduced information transfer. Similarly, Jacob *et al.* [199] propose a definition of heterogeneity based on the degree spectrum rather than degree values. Specifically, a graph is considered heterogeneous if all the nodes have a different degree. Conversely, a star network is considered strongly homogeneous because only one node has a different degree from all others. The study can be extended to the weighted case, by computing a weighted degree and analysing the in-degree and the out-degree distributions. Gao *et al.* [201] identify heterogeneity as a strong predictor of network resilience, with more heterogeneous networks being more resilient to targeted attacks but less resilient to random failures. Heterogeneity is computed over a directed weighted graph, according to the density functions of the in- and out-degree variables (k_{in} and k_{out}, respectively), *i.e.,* formally:

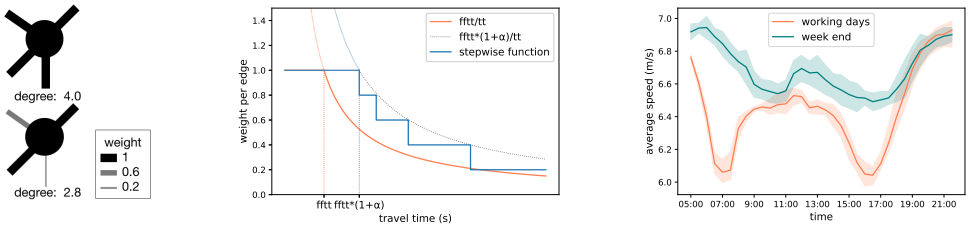
$$h = \frac{\sigma_{in}\sigma_{out}}{\langle k \rangle} \tag{4.6}$$

where σ^2_{in} (respectively, σ^2_{out}) is the variance of the density function of the weighted in-degree $P(k_{in})$ (respectively, out-degree $P(k_{out})$) and $\langle k \rangle$ is the average degree or network density, defined as follows:

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i \tag{4.7}$$

where N is the number of nodes in the network and k_i is the degree centrality of the node i, *i.e.,* $k_{in}(i) + k_{out}(i)$.

To quantify the resilience of different types of multi-dimensional systems, the authors propose to use a combination of three network metrics derived from the degree centrality,



(a) Node's unweighted (top) and weighted (bottom) degree centrality

(b) Edge weight computation at time slot t based on the (average) observed travel time

(c) The average speed profile (m/s) for working days and during the weekend.

Figure 4.4: The time-varying network modelling approach.

namely: heterogeneity h, density $\langle k \rangle$ and network symmetry S. Symmetry is defined as the correlation coefficient of in-degree and out-degree, *i.e.,* formally:

$$S = \frac{\langle k_{in} k_{out} \rangle - \langle k_{in} \rangle \langle k_{out} \rangle}{\sigma_{in} \sigma_{out}} \tag{4.8}$$

where $\langle k_{in} \rangle$ (respectively, $\langle k_{out} \rangle$) is the average in-degree (respectively, out-degree), $\langle k_{in} k_{out} \rangle$ the scalar product of both vectors (in-degree and out-degree).

We thus explored the possibility to apply the metrics proposed by Gao *et al.* in the context of road traffic networks, by studying their sensitivity to different disruptions and discovered the existence of relevant patterns, which could be exploited to quickly detect the appearance of disruptions in the network and anticipate recovery actions.

In the following, we provide a brief description of the methodology and the most relevant outcomes of its evaluation. The reader can refer to [198] for more details on the implementation, the dataset and the complete set of performed experiments.

Network Modelling Approach

We modelled the road network of the city of Lyon, France, as a weighted, time-varying, directed graph, $G(V, E, W^t)$, with nodes V corresponding to road intersections and edges E corresponding to road segments. Edge weights W^t were assumed to indicate dynamic traffic conditions (*i.e.,* average speeds), computed every 30 minutes from available GPS floating car data. The free-flow travel time was used when no observation was available on a given edge at a given time slot in the reference period.

The first step of the proposed methodology consists in building a graph from the road network topology of the analysed city. Edge weights, at time t, are computed in the range [0, 1] by converting the average speed observed in that specific time slot into a traffic congestion indicator, with 1 indicating the best traffic conditions and 0 indicating the worst (Fig. 4.4(a)). In free flow conditions, all nodes are assumed to be connected at the best possible level, and all edges receive a weight equal to one. When the observed average travel time is higher compared to the free-flow travel time (*fftt*), the edge weight becomes closer to zero, according to the ratio of *fftt* and the observed average travel time (*tt*) (*i.e.,* the orange curve in Fig. 4.4(b)), with zero corresponding to the case of a completely congested edge, *i.e.,* a road segment where vehicles are completely stuck. A discretization process was applied to the edge weights in order to model bounded rationality in travellers' route choice behaviour [202]. The idea is that a small travel time increase should produce a negligible impact on the edge weight (*i.e.,* the step-wise function in Fig. 4.4(b)).

The analysis reported below explores the impact of selected disruptions on the road network of Lyon, through the lens of heterogeneity (Eq. 4.6), density (Eq. 4.7) and symmetry

(Eq. 4.8). The metrics are computed at each time step of 30 minutes, via the proposed model, under both normal and disturbed conditions. The analysis aimed to prove that they are indeed very sensitive to modifications of the travel demand, minor perturbations and extreme weather conditions. The results of the analysis revealed clear patterns in how the road traffic network reacts to day-to-day disruptions and more extreme disturbances, as evidenced by variations in the aforementioned metrics with respect to typical values observed in normal conditions. These findings are consistent with previous studies on static networks in other domains, such as Gao *et al.* [201].

Experimental Results

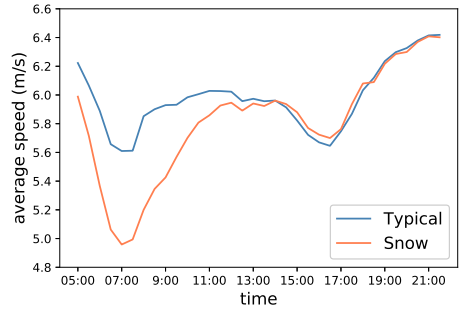
We performed an extensive evaluation of the previously described degree-centrality metrics in relation to four different disruptions that occurred in the period covered by the available GPS data[4]. The specific days were chosen based on the severity of the event or observed significant differences in speed or speed variance between disrupted and typical days.

On Monday, December 18$^{\text{th}}$, 2017, a heavy snowfall in Lyon caused severe disruptions in the road network, which lasted until the afternoon when traffic returned to normal. On Tuesday, December 19$^{\text{th}}$, 2017, the subway service was shut down for maintenance from 7:30 am to 4:15 pm. We studied the impact of this event on the road network because of the potential modal shift it generated. On Wednesday, April 4$^{\text{th}}$, 2018, several protests took place in Lyon, affecting normal traffic circulation on the road network. Finally, on Saturday, June 2$^{\text{nd}}$, 2018, the Fourvière tunnel, which is crossed by over 100,000 vehicles daily, was closed for renovations in the north-south direction for three days. Although some of these disruptions affected the entire network, others were localised in specific areas of the city. We report here results related to December 18$^{\text{th}}$ (heavy snowfall) and June 2$^{\text{nd}}$ (closed tunnel) and refer readers to our paper [198] for further details on the other days.

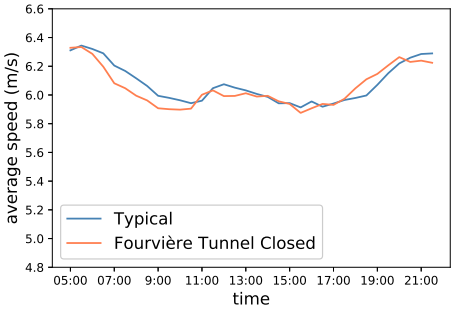
We observed significant variance of the aggregate speed variable computed on our dataset across the different days of the week (transparent margins in Fig. 4.4(c)). Therefore, we compared the speed profiles of disrupted days with the typical speed profile of the same day of the week (without major disruptions). For instance, if a disturbance occurred on a Monday, we compared the traffic conditions of that day with a typical (non-disrupted) Monday. A similar approach was used to study the time profile of the average degree (*i.e.,* network density).

We begin by examining the typical speed profiles for vehicles on the same weekdays when the snowfall and closed tunnel events occurred. These profiles are represented by the blue curves in Fig. 4.5(a) for the snowfall scenario, and Fig. 4.5(b) for the closed tunnel scenario. Upon comparison, we find that the typical speed in the snowfall scenario (Fig. 4.5(a)) is lower than that in the closed tunnel scenario (Fig. 4.5(b)). This is because the first scenario occurs on a weekday (Monday), while the second one occurs on a weekend day (Saturday), with obviously different traffic conditions. The same trend appears in the typical degree profile displayed in blue (Figures 4.5(c) and 4.5(d)). The histograms in Figures 4.5(e) and 4.5(f) report the degree distributions, focusing on the morning peak time, 7:30 am. This is the time slot when we consistently noticed a significant gap between the typical average speed profiles and those related to disrupted events. For all studied cases (Figures 4.5(e) and 4.5(f)), we observe a higher number of nodes possessing a degree value between 6 and 9 in the typical distribution than in the disrupted one. However, for the ranges grouping small degree values below 3, the trend is reversed: there are more nodes for such degree values in the deteriorated situation than in the normal one. This result is not surprising and indicates that, in the presence of disruptions, the degree distributions shift towards lower values (worse traffic conditions). This is especially the case for the first disruption (Fig.

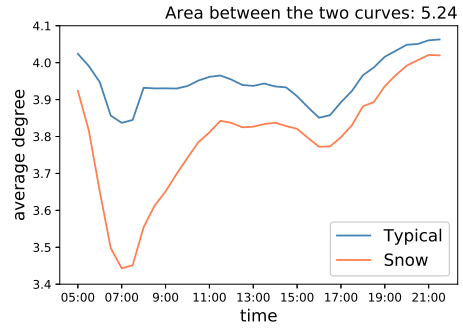
[4]The data, acquired by the FCD provider Media Mobile (now Be-Mobile), were available from October 2017 to September 2018 (see also Sec. 3.1, where the same data were used for speed estimation from UAPD data).



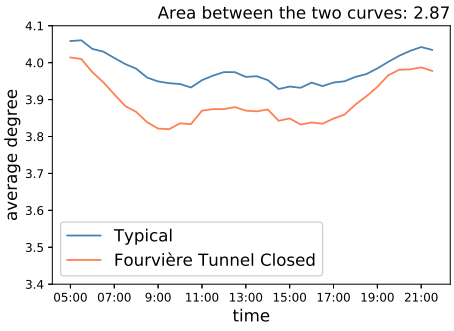
(a) Snow, Average Speed Profile



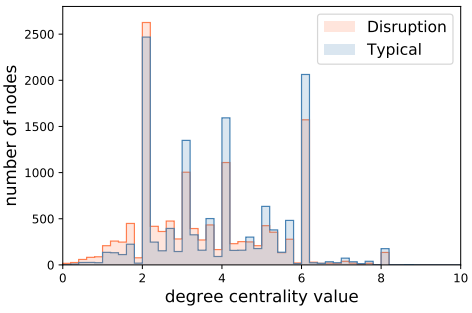
(b) Tunnel Closure, Average Speed Profile



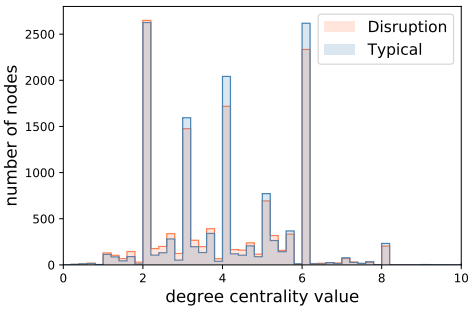
(c) Snow, Average Degree Profile



(d) Tunnel Closure, Average Degree Profile



(e) Snow, Degree Distribution at 7:30



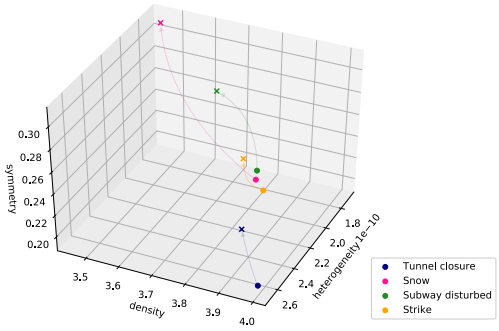
(f) Tunnel Closure, Degree Distribution at 7:30

Figure 4.5: Average speed profile, average degree centrality profile and degree distribution during two disruptive events: snow on December 18$^{\text{th}}$, 2017 (left) and tunnel closure on June 2$^{\text{nd}}$, 2018 (right).

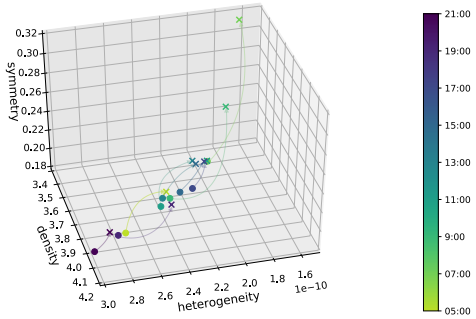
4.5(e)), which has the most significant impact on the average speed profile (Fig. 4.5(a)).

The comparison of the gap between the average degree centrality curves in the two disrupted scenarios confirms this result. Whereas the area between the curves is equal to 5.24 for the first disruption (Fig. 4.5(c)), it is only equal to 2.87 for the second one (Fig. 4.5(d)). In general, the average degree is always lower in disrupted conditions than in normal conditions on the same day of the week, showing a certain sensitivity of this global metric to events affecting the network. It is also worth highlighting that the gap is larger in the average degree centrality than in the average speed profiles, likely due to the weight discretization process, introduced to model bounded rationality in route choice.

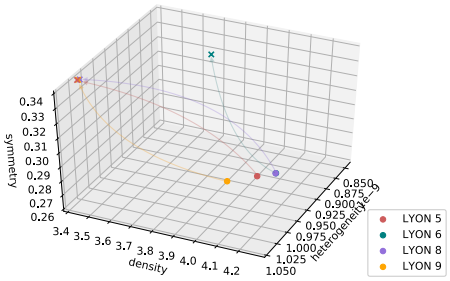
We now focus on analysing the patterns associated with the combination of network heterogeneity, symmetry and density, in both normal and disrupted conditions. Similarly to [201], we visualise the values of these metrics in a three-dimensional plot (Fig. 4.6) in



(a) Several disruptions at 7:30am



(b) December 18$^{\text{th}}$, temporal analysis on typical/disrupted moments of the day. Segments connect corresponding time slots on the different days.



(d) December 18$^{\text{th}}$, spatial analysis at 7:30am

Figure 4.6: Analysis of disruptions in the heterogeneity-symmetry-density space.

order to characterise the network under different traffic conditions.

In the presence of a disruption (symbol $\times$ in Fig. 4.6), the graph density and its heterogeneity decrease, while the symmetry increases compared to reference situations (symbol $\bullet$ in Fig. 4.6). The previous observations concerning the degree distribution in the presence of disruptions (*i.e.,* values concentrating in lower ranges than in normal conditions due to reduced speeds in the presence of disruptions) easily explain the reduction in network density. The decrease in heterogeneity (Eq. 4.6) can be instead explained via the lower observed values of the standard deviation of the in-degree and out-degree distributions. Finally, the increase in the symmetry value reflects an increase in the correlation coefficient of in-degree and out-degree distributions. When a disruption occurs, both in-degree and out-degree values become generally lower, thus reducing the corresponding average degrees. However, this reduction is accompanied by a stronger decrease in the standard deviation of the degrees, which explains the increase in symmetry. Remarkably, the impact of disruptions ($\times$) varies in intensity. As Fig. 4.6(a) highlights, the effect of snowfall (red symbols) is stronger than the other considered scenarios. Finally, the tunnel closure scenario (blue points and lines) is rather isolated in the plot space with respect to the other disruptions. The former occurred on a Saturday, while all others happened on weekdays when typical travel times are lower (Fig. 4.4(c)), explaining the difference in network states.

Our dynamic graph modelling approach also allows, differently from [201], to follow the evolution over time of the metrics in the three-dimensional plot. Specifically, we compared heterogeneity, symmetry and density every 30 minutes on a typical Monday and on the day (a Monday) of the heavy snow (Fig. 4.6(b)). The figure shows, with the same colour, but

with a different symbol, the triple $(h, \langle k \rangle, S)$ at each given time slot, for both the typical and the disrupted situations. It can be noted that, in all cases, the time slots on the disrupted day exhibit lower heterogeneity, lower density and higher or similar symmetry values, confirming the already observed global trend. This aspect denotes that the disruption had negative effects on traffic throughout the whole day but with increased intensity in specific moments of the day. Specifically, on a regular Monday, the worst traffic conditions are observed at 7:00 am, *i.e.,* during the morning peak hour (green • points in Fig. 4.6(b)), while good traffic conditions can be observed, *e.g.,* at 7:00 pm, in accordance with the speed profile of working days (Fig. 4.4(c)). When considering the day disrupted by the snowfall, the × points always exhibit higher symmetry, lower density and lower heterogeneity with respect to the regular day, independently of the time slot. The most significant changes between the typical state and the disrupted one can be observed during morning, *i.e.,* at 7:00 am and 9:00 am (Fig. 4.6(b)), corresponding to the moments when the snow had actually the most disruptive impact on the transport network.

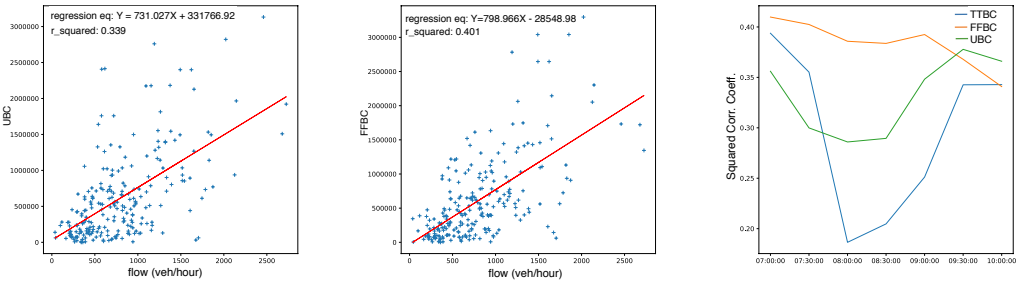
We also conducted a spatial analysis by separately computing the values of the three metrics in four districts of the city, at 7:30 am. Fig. 4.6(d) presents the per-district network density, heterogeneity, and symmetry on December 18^{th}. It is relevant to underline that all areas exhibit the usual pattern in the presence of disruption: density and heterogeneity decrease, while the symmetry increases. For the typical Monday, the network states are close, even overlapping, for the 6^{th} and the 8^{th} districts under normal conditions. During the snowfall instead, in all cases, an increase of symmetry can be observed, but a significant reduction of heterogeneity is only observed in the 6^{th} district, more urbanised and probably more affected by the snowfall. Similar considerations could be observed also for the other disruptive events, in both the local and global analyses.

In conclusion, the proposed approach utilising a dynamic, graph-based model and incorporating the concepts of heterogeneity, symmetry, and average degree revealed interesting insights into the behaviour of road traffic networks exposed to perturbations. The conducted experiments identified a specific pattern, where lower values of density and heterogeneity and higher values of symmetry appear to be associated with more significant deterioration of traffic conditions. This indicates a disrupted behaviour in contrast to typical baseline scenarios. Monitoring these topological indicators can thus provide valuable real-time information about a potential incoming perturbation, thus helping in preparing and recovering from it. Furthermore, these indicators can be used to compare different subsets of network nodes, such as different neighbourhoods within a city, to assess their robustness to disruptions.

4.2.2 Correlation of Betweenness Centrality and Traffic Metrics

In our efforts to explore useful metrics for monitoring transport networks, we investigated in [203] the properties of edge BC (Eq. 4.1) when computed on dynamically weighted graphs. We followed a data-driven modelling approach similar to the one described in Sec. 4.2.1 but with a different weighting procedure. Instead of using a congestion indicator, we assigned the average travel time to each edge, periodically computed every 30 minutes. We chose this approach to be able to determine temporal shortest paths in the definition of BC. In a first case study, we computed average travel times from real GPS FCD, and in the second study, we retrieved them through simulations using realistic data on demand and supply for the city of Lyon, France. In the following, we report the main outcomes related to the first case study. More details can be found in [203].

Our analysis confirmed the existence of a non-negligible correlation between BC and traffic flow, as already observed in related literature [204, 205]. In addition, we explored and clarified the nature of such correlation from a spatio-temporal perspective, as well as a function of structural and behavioural traffic properties. Similarly to the topological static approach discussed in Sec. 4.1.1, we considered different variations of edge BC, namely, *(i)* UBC, in which the graph is considered static and unweighted; *(ii)* Free-Flow weighted Be-



(a) UBC ($r^2 = 0.34$) (b) FFBC ($r^2 = 0.40$) (c) Evolution of r^2 over time

Figure 4.7: Correlation of per-edge average traffic flow and different BC metrics (UBC, FFBC, TTBC).

(a) Per-edge temporal correlation (only edges equipped with loop detectors have a non-null value)

(b) Zoom on an area with two roads: Qaui Dr. Gailleton and Quai Claude Bernard

(c) Evolution of the TTBC and flow on the road: Qaui Dr. Gailleton and Quai Claude Bernard

Figure 4.8: Per-edge temporal correlation between BC and flow (a), and zoom on a specific region (b,c).

tweenness Centrality (FFBC), computed considering static free-flow travel times as weights on each edge; *(iii)* TTBC, which uses our dynamic data-driven approach with travel time information computed every 30 minutes and associated to the edges as weights.

In Fig. 4.7(a) and Fig. 4.7(b), we present the values of what we called *spatial correlation*. The latter was obtained by performing a linear least squares regression between the selected edge BC metrics (computed on a static graph) and the flows observed via loop detectors on the corresponding network edges. Each point corresponds to an edge with associated values of hourly flow (on the x-axis) and the edge's value of BC (on the y-axis). The results show a 0.34 value of r^2 for UBC and a 0.40 value of r^2 for FFBC, pointing to some mild capability of static BC metrics to explain flow distribution on the network.

We also found that TTBC is sensitive to traffic flow dynamics when computed on a time-varying weighted graph with average travel time weights: the correlation of edge BC and traffic flow varies, rather significantly, over time, decreasing during peak hours associated to higher demand, and, consequently, congestion. Fig. 4.7(c) shows the evolution over the different time slots of the r^2 coefficient of the linear regression between flow and three metrics: the static (UBC and FFBC) and the dynamic (TTBC) ones, the latter computed using average taxi travel times. TTBC appears to be the most sensitive to traffic dynamics among the three different metrics. Indeed, we observe an important drop of correlation during rush hours 07:30-09:00 for TTBC whereas the correlation for the FFBC metric remains practically constant, with mild variations only due to the dynamics of the traffic flow.

This phenomenon was partially explained by the existence of anti-correlation patterns

(*i.e.,* reduction of BC in the presence of increased traffic flow, and vice-versa) for some edges of the analysed network, which was confirmed both in simulated and real-world experiments. To identify these patterns, we studied what we called *per-edge temporal correlation*. This correlation was computed by considering, for a given edge l of the graph, the vector corresponding to the TTBC values over the available time slots and the vector of the corresponding observed vehicle flows (measured as vehicles/hour). Hence, we calculated the linear correlation score (which can be either positive or negative) for each edge l of the network. The edge temporal correlation values are spatially visualised in Fig. 4.8(a). In the figure, highly anti-correlated edges can be spotted (orange/red links) from the positively-correlated ones (blue/green links), while yellow links correspond to weakly-correlated edges.

We analysed the dynamics related to TTBC and flows by focusing on a specific region of the graph, including two roads with a mirror behaviour in terms of temporal correlation, *i.e.,* Quai Dr. Gailleton (QDG) and Quai Claude Bernard (QCB). Since the FFBC of QDG is much higher than the FFBC of QCB (see values reported in Fig. 4.8(b)), QDG should typically attract a larger flow than QCB, as confirmed by the observed average flow of QDG equals to 2,679 veh/hour, significantly higher than the one observed on QCB (967 vehicles/hour). A high value of FFBC combined with a high average flow suggests that the link can be often expected to be close to critical flow. We confirmed from the available speed data that congestion is actually observed on QDG at two different time slots, *i.e.,* 08:00 and 08:30. Our data also indicates that when congestion happens on this link, an increase in travel time is also observed on the link. This is coupled to a decrease of TTBC, which thus exhibits an anti-correlation behaviour with respect to traffic flow. Conversely, the QCB link, which is characterised by lower demand, appears to be in free-flow conditions during the whole observation period. As travel time is rather high for QDG during congestion, QCB becomes then a viable alternative for drivers to avoid congestion on QDG. Consequently, we observe that TTBC increases as well as its flow. This explains why the QCB edge exhibits a highly positive temporal correlation. After the congestion phase (08:00 and 08:30), the flow globally decreases in the area. As QDG becomes more fluid, the corresponding travel time decreases to free-flow travel time and TTBC increases, *i.e.,* thus maintaining an anti-correlated tendency. Contextually, QCB appears to lose its attractiveness compared to QDG as the flow decreases on this edge. TTBC decreases as well thus keeping the positively correlated trend. The analysis is confirmed by the evolution of the flow and the TTBC, which are known from the available data for both QDG and QCB as reported in Fig. 4.8(c).

To summarise, FFBC gives information on critical edges, *i.e.,* road segments on which a high flow and possible congestion should be expected. Then, by studying the dynamic behaviour of TTBC, we highlighted different types of edges. On the one hand, anti-correlated edges seem to be critical by nature with higher FFBC. Even when congestion is observed on these links, the flow tends to remain high, whereas TTBC decreases. On the other hand, positively correlated edges appear to correspond to alternative routes in terms of travel time that become attractive when nearby roads are congested. In an intermediate situation, *neutral* edges (low correlation) are characterised by either low or medium/high demand, being able to dispatch such flow, without becoming congested.

Overall, the results of this study let us speculate that, if *(i)* people exploit the knowledge about current traffic conditions, *(ii)* alternative paths are available, and *(iii)* no specific constraint obliges travellers to follow specific routes, vehicles can re-distribute themselves in the network more efficiently and better exploit it. In this particular case, it can be assumed that edges with higher BC in a certain time slot represent critical (or very attractive) components of the network where traffic flow will most likely significantly increase in the near future. These properties could thus justify using this metric for the implementation of next-generation proactive, data-driven urban monitoring systems that can use alternative route recommendations to reduce traffic congestion and vulnerabilities.

These conclusions led us to investigate: *(i)* the possibility of computing in real-time, on weighted time-varying graphs, BC on large networks, which is a very challenging task due

to the high processing time of state-of-the-art BC computation algorithms; *(ii)* considering traffic control solutions that could take into account real-time information on traffic flow and BC to protect specific areas of the network, expected to attract massive traffic in the future. The research related to these two problems related to real-time monitoring and control is described in the next chapter. The following section focuses instead on managing disruptive situations occurring with a known (relatively low) frequency via multi-modal transport solutions for robust and sustainable network design.

4.3 Vulnerability-aware Multi-modal Network Design

The knowledge of a transport network's vulnerabilities, acquired, *e.g.,* with the tools introduced in Sec. 4.1, can be harnessed in decision-making systems to enhance the network's ability to respond to specific disruptions. This corresponds to the *knowing what to do* cornerstone of resilience, as illustrated in Fig. 1.4(b). Within this research scope, we have explored *on-demand and multi-modal mobility services* as a means *(i)* to ensure greater robustness in the face of atypical events and *(ii)* to improve adaptability to large variations in travel demand that have already occurred.

Concerning the first point, *i.e.,* increasing robustness to different kinds of disruptive scenarios, we have explored, in collaboration with the University of New South Wales, Sidney, Australia, the integration of *on-demand* mobility solutions into an existing large-scale urban transport network by taking into account the occurrence probability and expected traffic dynamics of a multitude of both regular and atypical scenarios [206]. In the same scope of research, we have also proposed using metrics for multi-layer networks to schedule the construction of new transit lines a transport operator might want to deploy to enrich its mobility offer [207]. Such methodologies typically fall within the broad field of transit network design. The novelty of these research activities [206, 207] lies in the consideration of multi-modality as a way for improving transport resilience. As an additional contribution, we considered the requirement of quick computation to allow for a rapid re-configuration of the transport supply, towards possible applications even in emergency situations.

Regarding the second point, *i.e.,* enhancing adaptability in the presence of extreme variations of travel demand, we have explored, in collaboration with Ghent University, Belgium, optimisation and machine-learning-based solutions [208] for adapting transit supply to the decreased use of public transportation. The methodology was tested in the COVID-19 pandemic scenario, during which lock-downs and safety-related restrictions led to an over-capacitated and less efficient public transport system.

In the following, we summarise the methodological contribution and the main results concerning the optimal deployment of an on-demand mobility service to increase transport resilience [206]. The other mentioned contributions are left to be explored by the interested reader in papers [207, 208].

4.3.1 Park-and-Ride for Resilient On-Demand Urban Mobility

In [206], we focused on designing a specific form of on-demand ride-sharing service, namely, Park-and-Ride (PR), to improve the resilience of an urban transport network with respect to disruptive scenarios.

PR systems are transportation services enabling people to drive their private vehicles to designated *pick-up* locations, usually situated in the outskirts of a city. These pick-ups have parking lots where travellers can commute to public or private ride-sharing modes such as buses, subway, or on-demand shuttles. People can use these modes to move to *drop-off* locations within the city and reach close-by destinations. The idea is to allow people to reduce the costs of their trips and decrease the number of private vehicles circulating in the urban area. In opposition to traditional fixed-schedule/fixed-routes transit systems, ride-

sharing services offer flexibility in terms of routes, departure/arrival time, fleet size, *etc.*, which can be helpful to reduce the total demand for car trips and increase city resilience [209].

Our study considered a PR facility location problem that captures users' entire trip chains – including access, transit and egress trips – in an urban mobility network. The PR system is assumed to comprise a shuttle service that complements the existing mobility network. User mobility outside of our PR system is captured through the *reserve mode*. In designing the PR facility, both typical mobility conditions and recurrent disruptive scenarios are taken into account. The objective is to maximise the overall ridership of the proposed PR system across all considered scenarios. Disruptions are modelled as scenarios with predetermined probabilities of occurrence, which must be known in advance. The aim was to specifically account for weather-related events (such as heavy rain and snowfall), pollution peaks, and other special events that could adversely affect regular transportation. The frequency of these disruptive events can be estimated in advance for a given urban area, in relation to typical traffic conditions. In our study, we quantified the attractiveness of the PR system by incorporating a simple model of user mode-choice behaviours into the formulation. This model is based on a multinomial logit choice model and seeks to maximise the expected ridership within the PR system at the expense of users' reserve travel option.

The proposed methodology provides: *(i)* the optimal location of pick-ups, *i.e.,* the car parks where commuters can transfer from their vehicle to shared mobility solutions (*e.g.,* shuttle buses); *(ii)* the optimal location of drop-offs, *i.e.,* the transit stops where users exit the PR system and continue to their destination by other means, such as walking, micro-mobility (*e.g.,* bike-sharing, e-scooters) solutions or the regular public transportation; and *(iii)* the expected flow share associated to the PR pick-up/drop-off pairs from the set of OD pairs, in each considered scenario. We assume a given budget constraint for constructing the car parks and transit stops.

Given the varying and hardly predictable nature of perturbations, as well as the large size and complexity of typical metropolitan mobility networks, guaranteeing a low computation time in solving the optimisation problem was considered a critical requirement to satisfy. By quickly solving the optimisation problem whenever a novel set/probability distribution of disturbances is identified, a different configuration of the park-and-ride system can be dynamically enacted (*e.g.,* changing the location of the drop-offs) by transport authorities.

The proposed solution builds on and significantly extends the approach of Aros-Vera *et al.* [210], where the authors only focused on the pick-up and transit part of users' trip, thus ignoring the impact of drop-off location with respect to the final destination.

The main contributions of our study are: *(i)* the mathematical formulation of an integrated PR system, which captures the entire user trip chain from the origin to the destination via pick-up and drop-off nodes in a mobility network and accounting for mode choice; *(ii)* the incorporation of a stochastic programming approach to take into account recurrent disruptions in an urban mobility network; *(iii)* the development of a customised Lagrangian Relaxation Algorithm (LRA) able to provide competitive solutions, by out-performing commercial Mixed-Integer Linear Programming (MILP) solver CPLEX, based on a branch-and-cut search, for large-scale mobility networks in a restricted computational time; *(iv)* the implementation of the proposed PR system on a realistic instance representing the city of Lyon, France, which provides key insights for mobility service providers in urban areas. In the following, we briefly detail points *(i)* and *(iv)*, leaving the interested reader to paper [206] for more details on the remaining contributions.

Mathematical formulation of an integrated PR system

We modelled the PR problem as a stochastic problem which aims at considering a set Ω of different scenarios ω, with a probability of occurrence p^{ω}, representing both normal conditions and recurrent disruptions, to optimise our PR system in an existing transport network. The stochastic nature of the problem allows making resilient PR design, by tak-

ing into account the probability of occurrence of multiple possible scenarios of disruptions. Scenario-dependent variables and parameters are denoted with a superscript $^{\omega}$ in the following. The transport network is represented by a directed graph with $|E|$ edges representing the roads and $|N|$ nodes corresponding to the road intersections. In each scenario, we determined the set of the OD pairs (r, s) with a non-null total travel demand d^{ω}_{rs}. This set is called $W^{\omega} = \{(r, s) : d^{\omega}_{rs} > 0, r \in N, s \in N\}$.

For each OD pair, it is possible to identify the sets of the accessible pick-ups i and drop-offs j. To be attractive, the generic PR pick-up i has to be reached from the origin r in a reasonable time t^{ω}_{ri} lower than a fixed time t^{access}. Similarly, the destination s must be reached from the drop-off j in a given time t^{ω}_{js} lower than the fixed egress time t^{egress}. The travel times t^{ω}_{ri} (respectively, t^{ω}_{js}) correspond to the shortest travel time between the origin r and the pick-up i of our road network (respectively, the drop-off j and the destination s). Ultimately, we introduced the set of the potential pick-ups $P^{\omega}_{r} = \{i \in P : t^{\omega}_{ri} \leq t^{\text{access}}\}$ among all the pick-ups $P \subseteq N$ and the set of potential drop-offs $D^{\omega}_{s} = \{j \in D : t^{\omega}_{js} \leq t^{\text{egress}}\}$ among all the drop-offs $D \subseteq N$. From these potential pick-ups and drop-offs, we extracted for each OD pair the set $\Sigma^{\omega}_{rs} = \{(i, j) : i \in P^{\omega}_{r}, j \in D^{\omega}_{s}, (r, s) \in W^{\omega}\}$ of the PR alternatives which correspond to the possible combinations of pick-ups and drop-offs at each OD pair in each scenario ω.

We defined a binary decision variable $y_i \in \{0, 1\}$, $\forall i \in P \cup D$, indicating the facility locations by being equal to one when a car park (pick-up) or a transit stop (drop-off) is open. Furthermore, we defined the decision variable $x^{\omega}_{rijs} \in \mathbb{R}^{+}, \quad \forall \omega \in \Omega, \forall (r, s) \in W^{\omega}, \forall (i, j) \in \Sigma^{\omega}_{rs} \cup \{\mathsf{R}_{rs}\}$, as the flow which determines the part of the scenario-dependent demand d^{ω}_{rs} associated to our PR system on the OD pair (r, s) using the pick-up i and the drop-off j. The decision variable $x^{\omega,\mathsf{R}_{rs}}_{rs}$ defines instead the part of the users that choose the reserve mode on the OD pair (r, s).

To determine the mode choice via a market share model based on a logit formulation, we used, for each itinerary, a generalised cost g^{ω}_{rijs} composed of the travel cost between r and s using the proposed PR system, that we considered equal to the travel time, between the three parts of the path: *(i)* from the origin r to the pick-up i, *i.e.,* $c^{\text{access},\omega}_{ri}$; *(ii)* from the pick-up i to the drop-off j, *i.e.,* $c^{\text{route},\omega}_{ij}$; *(iii)* and from the drop-off j to the destination s, *i.e.,* c^{walk}_{js}. When using the reserve mode to realise the entire itinerary, symbolised as R_{rs}, the travel cost $c^{\mathsf{R}_{rs},\omega}_{rs}$, equal to the generalised cost $g^{\mathsf{R}_{rs}}_{rs}$, is modelled as the travel time to reach the destination s from the origin r by car. The probability that users choose the pair i and j of the PR system corresponds to the x^{ω}_{rijs} decision variable, which can be expressed as follows:

$$x^{\omega}_{rijs} \equiv \frac{y_i y_j e^{-\theta g^{\omega}_{rijs}}}{\sum\limits_{(m,n) \in \Sigma^{\omega}_{rs} \cup \{\mathsf{R}_{rs}\}} y_m y_n e^{-\theta g^{\omega}_{rmns}}} \quad \forall (r, s) \in W, \forall (i, j) \in \Sigma^{\omega}_{rs} \cup \{\mathsf{R}_{rs}\}. \tag{4.9}$$

In other words, variable x^{ω}_{rijs} represents the probability of using the reserve mode between the OD pair (r, s), when i and j are equal to R_{rs}, as well as the probability of using the proposed PR system by using a shuttle joining the pick up i to the drop off j for the OD pair (r, s). Because both pick up and drop off have to be open to be considered as an option in the mode choice, the exponential of the utility for all the transport mode possibilities are multiplied by the pick up and the drop off location decision variables y_i and y_j. The closure of a pick up, a drop off or both will remove the consideration of the associated transport modes in the logit computation. We denoted θ the logit parameter corresponding to the user's sensitivity to the generalised cost. The higher θ, the more people are sensitive to the travel cost and will choose the cheaper transport mode. Finally, $-\theta g^{\mathsf{R}_{rs},\omega}_{rs}$ (*i.e.,* $g^{\mathsf{R}_{rs},\omega}_{rs} = g^{\omega}_{rijs}$ with i and j equal to R_{rs}) corresponds to the utility of the reserve mode and $-\theta g^{\omega}_{rijs}$ represents the utility of the PR system from the pick-up i to the drop-off j for the path going from the origin r to the destination s. Although simple, our utility function considers the travel time preference in the mode choice. A more complex function could more accurately

represent this mode choice, for instance, by being trip purpose-dependent [211].

A budget B is allowed to the construction of the pick-ups and the drop-offs whose costs $\boldsymbol{c}_i^{\text{loc}}$ mostly depend on the nature of the facility (Eq. 4.10b). Whereas car parks must be constructed at the pick-up locations for users to park their cars and embark on the shuttles of the PR system, transit stops are sufficient to leave the shuttle and reach the final destination by walking.

The resulting PR facility location formulation summarised in (4.10) is a MILP.

$$\max \quad \sum_{\omega\in\Omega} p^{\omega} \sum_{(r,s)\in W^{\omega}} d_{rs}^{\omega} \sum_{(i,j)\in\Sigma_{rs}^{\omega}} x_{rijs}^{\omega} \tag{4.10a}$$

$$\text{s.t.} \sum_{i\in P\cup D} c_i^{\text{loc}} y_i \leq B \tag{4.10b}$$

$$\sum_{(i,j)\in\Sigma_{rs}^{\omega}\cup\{\mathsf{R}_{rs}\}} x_{rijs}^{\omega} = 1 \qquad \forall\omega\in\Omega, \forall(r,s)\in W^{\omega} \tag{4.10c}$$

$$x_{rijs}^{\omega} \leq x_{rmns}^{\omega} \frac{e^{-\theta g_{rijs}^{\omega}}}{e^{-\theta g_{rmns}^{\omega}}} + (2 - y_m - y_n) \qquad \forall\omega\in\Omega, \forall(r,s)\in W^{\omega}, \forall(m,n),(i,j)\in\Sigma_{rs}^{\omega}\cup\{\mathsf{R}_{rs}\} : (i,j)\neq(m,n) \tag{4.10d}$$

$$x_{rijs}^{\omega} \leq y_i \qquad \forall\omega\in\Omega, \forall(r,s)\in W^{\omega}, \forall(i,j)\in\Sigma_{rs}^{\omega} \tag{4.10e}$$

$$x_{rijs}^{\omega} \leq y_j \qquad \forall\omega\in\Omega, \forall(r,s)\in W^{\omega}, \forall(i,j)\in\Sigma_{rs}^{\omega} \tag{4.10f}$$

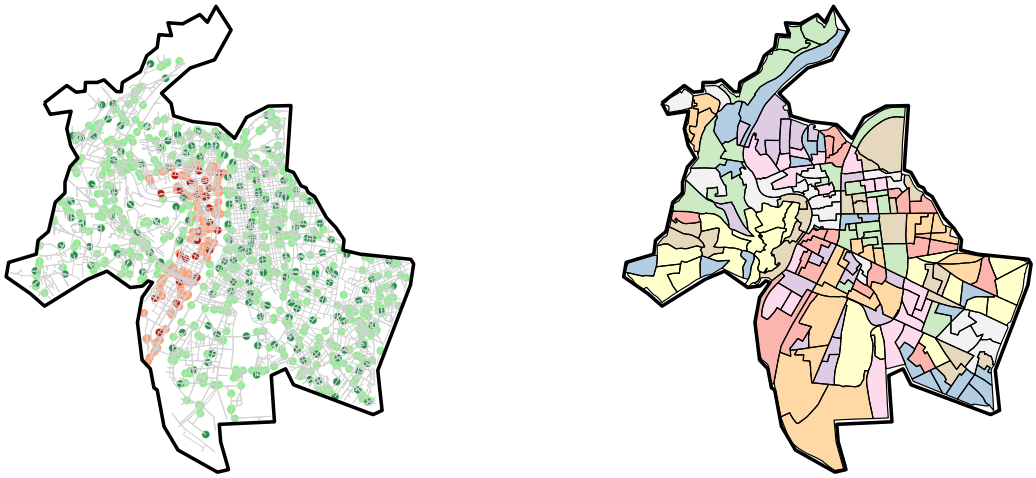
$$y_{\mathsf{R}_{rs}} = 1 \qquad \forall\omega\in\Omega, \forall(r,s)\in W^{\omega} \tag{4.10g}$$

$$x_{rijs}^{\omega} \geq 0 \qquad \forall\omega\in\Omega, \forall(r,s)\in W^{\omega}, \forall(i,j)\in\Sigma_{rs}^{\omega} \tag{4.10h}$$

$$y_i \in \{0,1\} \qquad \forall i\in P\cup D \tag{4.10i}$$

By maximising the objective function (4.10a), the optimal solution will maximise the portion x_{rijs}^{ω} of the demand d_{rs}^{ω} using our PR system in each scenario ω. The higher the probability p^{ω} related to the occurrence of a given scenario ω, the higher the weight this scenario has in the objective function. The first constraint (4.10b) determines the ability to build the facilities by satisfying the global cost of the open parks. Constraint (4.10c) ensures that the whole demand is served for each origin-destination pair. Users must perform these trips through the PR system or using the reserve mode. The second constraint (4.10d) requires that the flow share of each mobility alternative obeys a linearised logit model. In [206], we provided the mathematical proofs, based on the ones proposed by [210], that lead to Eq. 4.10d from Eq. 4.9, which controls the proportion of the users for each mobility alternative. For each OD pair, this choice is governed by the utility of mobility alternatives through the PR system, which is a function of opened PR facilities and of the reserve mode. For a given PR path (r,i,j,s), if the pick-up location y_i or the drop-off location y_j are closed, the constraint is inactive due to linking constraints (4.10e) and (4.10f), forcing the flow share to be lower than (or equal to) one when parks are open, and zero otherwise. For a given OD pair (r,s), it is only possible to have a positive share of travellers using path (r,i,j,s) if and only if $y_i = y_j = 1$. The decision variable representing the portion of users going from the origin r to the destination s through the pick-up i and the drop-off j, x_{rijs}^{ω}, must be positive (4.10h). Constraints (4.10g) and (4.10i) set the domain of binary variables y_i and fix all such variables to one for the mobility alternative corresponding to the reserve mode.

The complexity of the problem highly increases with the graph size due to the number of variables. For instance, with a small network composed of 59 nodes used in our numerical experiments, the formulation corresponds to 267 000 constraints and 15 000 variables; the problem grows to 15 000 000 constraints and 270 000 variables for a larger one, composed of



(a) Locations of the origins (light green), destinations (light red), potential pick-ups (dark green) and potential drop-offs (dark red)

(b) Iris areas: territory divisions encompassing 2 000 inhabitants

Figure 4.9: Network segmentation (a); and the network used for the real instance I^R (b).

135 nodes. These figures made the problem hard or impossible to solve with commercial solvers and led us to develop a decomposition approach based on Lagrangian Relaxation (LR). The proposed LR, widely discussed in [206], provides upper bounds on the original problem and uses customised heuristic algorithms to iteratively generate lower bounds during the solution of the LR problem.

Experimental Results in a Realistic Case Study

The proposed formulation was solved in different case studies with both a commercial CPLEX solver and our LRA. All of the case studies were based on the road network of Lyon, France, composed of 10 905 nodes and 19 703 edges. The geography of the city, which is characterised by a peninsula surrounded by two rivers, makes the city centre particularly exposed to traffic disruptions and weather events, due to the limited number of alternative routes. The city centre is also the location of major cultural events, tourist attractions, and commercial establishments that attract large numbers of people. Finally, the pedestrianisation of Lyon's peninsula is currently an active topic of discussion in the city. For these reasons, we investigated the potential of an alternative, on-demand transport mode to improve accessibility and reduce the use of private vehicles. We focused on trips from outside the peninsula to the peninsula. All trip origins (r) and potential locations of pick-ups (i) were placed outside the peninsula, while all potential locations of drop-offs (j) and destinations (s) were considered within the first and second neighbourhoods that make up the peninsula.

Travel costs were determined using the FCD from October 2017 to September 2018, already introduced in Sections 4.2.1 and 3.1. The observed travel times were used to compute travel costs between each origin and potential pick-up locations (*i.e.,* $c_{ri}^{\text{access},\omega}$ costs) and between each pick-up and potential drop-off locations (*i.e.,* $c_{ij}^{\text{route},\omega}$ costs). The travel costs between the drop-offs and the destinations (*i.e.,* c_{js}^{walk}) were computed by assuming a walking speed of $1.5m/s$ over the shortest paths of the road network. The travel costs associated with the reserve mode (*i.e.,* $c_{rs}^{\text{R}_{rs},\omega}$) were calculated as the average observed travel times along the shortest path for each OD pair. For the sake of simplicity, we ignored any additional cost associated with the reserve transport mode (*e.g.,* fuel, parking, congestion, pollution).

Instance	Parameters			MILP		LRA		
	t^{access}	t^{egress}	Budget (%)	Objective	Gap (%)	LB	Gap (%)	Gain (%)
I^R	2	8	5	287.956	2.9	902.346	51.5	68.1
			10	1 156.906	0.3	1 352.844	27.3	14.4
			15	1 408.596	0.2	1 534.226	17.6	8.1
			20	1 618.571	0.0	1 618.593	13.0	0.0

Table 4.4: Results on I^R after 1 hour, with a heuristic time constraint of 20 minutes.

In the realistic case study, we considered four scenarios: *i)* a snowfall, *ii)* a public transport disruption with two subway lines stopped, *iii)* a typical weekday, and *iv)* a typical weekend day (see Sec. 4.2.1). Each scenario presents a different speed profile representative of the traffic dynamics during the corresponding event. Regardless of the studied scenario, we weighted the graph using the travel times recorded at 7:30 am, when the impacts of disruptions were the worst. The occurrence probability of each scenario was computed as follows. The snowfall was considered to have an impact similar to rainfall and we fixed its occurrence to 35%. The subway disruption scenario was considered exemplary of a public transport failure, representing a rather recurrent event. We fixed its probability to 30%. Finally, the probability for the normal weekday (respectively, weekend day) was set to 25% (respectively, 10%). Regarding the demand d^{ω}_{rs}, we used realistic data reconstructed by other members of our laboratory [212] by relying on a methodology combining survey-based information, simulations and measures of flows collected via loop detectors installed in the city of Lyon.

In our experiments, we assumed that the construction cost c^{loc}_i of pick-up nodes is ten times higher than that of drop-off nodes, as the former typically involves constructing a car park while the latter may only require a transit stop. The budget allocated to facility location construction, denoted as B, was defined as a percentage of the total cost for opening all facility locations. In all our analyses, we considered four different budgets corresponding to 20%, 40%, 60%, or 80% of the total construction cost. The trip generalised costs g^{ω}_{rijs} were determined by considering the travel costs $c^{\mathsf{R}_{rs},\omega}_{rs}$, $c^{\text{access},\omega}_{ri}$, $c^{\text{route},\omega}_{ij}$, c^{walk}_{js}, as well as the demand d^{ω}_{rs} and the construction costs of pick-up and drop-off nodes c^{loc}_i. We conducted multiple sensitivity analyses (reported in [206]) to compare the performance of the proposed LRA approach with a direct MILP approach. Finally, only the morning peak hour was considered in this study to configure the PR system, by considering that the role of pick-ups and drop-offs would be reversed during the evening commute.

We assumed that the set of origins and destinations differs from the set of potential pick-up and drop-off nodes (as shown in Fig. 4.9(a)). To distribute the potential pick-up and drop-off nodes uniformly, we employed a segmentation of the city called Iris sectors (as shown in Fig. 4.9(b)). These sectors were developed by the French Institute of Statistics to divide the conurbation of Lyon into small geographical areas, each containing approximately 2,000 inhabitants[5]. We assumed that only one pick-up or drop-off node could be placed in each Iris sector. The location of the potential pick-up or drop-off was chosen as the road network node closest to the Iris area centroid. As for origin and destination nodes, we selected them to correspond to one-tenth of the road intersection nodes in Lyon's road network and uniformly distributed them throughout the city network. The final instance I^R included 745 origins and 171 destinations with 155 potential pick-ups and 25 potential drop-offs. The access and egress times were respectively fixed at $t^{\text{access}} = 2$ minutes and $t^{\text{egress}} = 8$ minutes. Such a problem corresponds to about 9 062 000 variables and induced around 77 000 constraints, according to Formulation 4.10[6].

[5] https://www.insee.fr/fr/metadonnees/definition/c1523

[6] The proposed formulations and algorithms were implemented in Python on a machine with 16 Gb of RAM and a CPU of 4.20 GHz (Intel(R) Core(TM) i7-7700K). All mixed-integer and/or linear programs were solved using CPLEX's Python API.

Scenario		Park-and-ride flow share			
Description	Prob.	Budget 5%	Budget 10%	Budget 15%	Budget 20%
Snowfall/Heavy Rainfall	0.35	42.67%	67.41%	76.84%	86.35%
Public transport disruption	0.30	46.95%	75.78%	82.70%	86.90%
Week day	0.25	54.82%	72.05%	86.25%	88.98%
Week-end	0.1	44.33%	69.56%	78.65%	82.79%

Table 4.5: Park-and-ride flow shares with different budgets, using $\theta = 0.1$, $t^{\text{access}} = 2$ minutes, and $t^{\text{egress}} = 8$ minutes.

Tab. 4.4 summarises the results obtained with the realistic instance. The sensitivity parameter to the generalised cost θ was chosen equal to 0.1. Whatever the allocated budget (5%, 10%, 15% and 20%), the LRA always outperforms the MILP approach, by always finding a feasible solution with a higher value of the objective function. For the smallest budget, the gain between both solutions even exceeds 50%. Although the gap of the LRA is important due to the small reduction of the upper bound, the lower bound, representing the best found feasible solution, is very interesting to solve the complex problem of allocating PR, respecting a market share model, on a large-scale network.

Table 4.5 summarises the flow shares using the PR system, depending on the allocated budget for the facilities construction and the scenario. The higher the budget, the more people use the PR due to the increase of possible path alternatives induced by the increase in the number of open pick-ups and drop-offs (graphically reported with the related budget in Fig. 4.10). The larger set of alternatives reduces the travel cost of the PR mode by locating pick-ups and drop-offs closer to origins and destinations. It can be noticed how the PR can properly handle recurrent disruptions, by being highly attractive even in the presence of snowfall or public transport disruption. Although the flow share is not the same for all the scenarios, the order of magnitude remains the same. Additional experiments also proved that an increase in access and egress times generally leads to increased attractiveness of the PR option by serving a larger part of the demand for the same allocated budgets. Increasing the access and egress times makes PR alternatives accessible to more origins and destinations, thus augmenting the part of the demand served with the same amount of opened pick-ups and drop-offs. In these configurations with higher values of access and egress time, the MILP approach could not provide a solution contrary to our LRA, proving again the interest of our proposed methodology.

In conclusion, the results obtained on the realistic instance proved the ability of the proposed LRA to provide information about the implementation of a budget-constrained optimised PR system, adapted to recurrent disruptions in a large-scale network. The results also provided insights about the flow share distribution between the reserve mode and the PR for a given allocated budget. While flow shares of this magnitude may be unrealistically high, this can be explained by the fact that only a single concurrent mobility mode was considered – the reserve mode – which may not be sufficient to fully capture competition effects across travel options in an urban multi-modal transportation network. This study has focused on the design of PR systems with the aim to improve the resilience of urban mobility networks in the presence of recurrent disruptions. However, the proposed facility location approach can be generalised beyond PR systems to the design of flexible multi-modal and on-demand mobility networks. Indeed, the proposed framework could be used to model three-link trip chains in a generic context. We are currently working to incorporate fleet sizing and vehicle or line capacity in the proposed framework. As another limitation of the approach, it is worth noting that the addition of the PR system may influence transit flows and subsequently traffic congestion in the network. The incorporation of congestion effects in the proposed framework may be achieved by including a traffic and/or transit assignment module in our formulations. Such extensions of the proposed formulation will be explored in future works.

(a) 5%: 12 P and 10 D

(b) 10%: 20 P and 15 D

(c) 15%: 29 P and 20 D

(d) 20%: 37 P and 20 D

Figure 4.10: Solutions found for the real instance: green nodes are origins (potential pick-ups) and red nodes are the destinations (potential drop-offs).

Chapter 5

Real-time and Large-scale Solutions for Resilient Urban Networks

As discussed in previous chapters, the resilience of transportation networks can significantly benefit from the availability of extensive, multi-source human mobility data. The latter typically arise from interactions between individuals and various communication devices, sensors, and applications. These data are stored and curated for later use in offline settings by traffic operators, telecommunication companies, and other service providers. As highlighted in Chapters 3 and 4, such data not only deepen our understanding of how transport networks function but also enable advanced strategies for vulnerability assessment and resilience-oriented design.

Modern cities have augmented this capability by generating real-time data streams through IoT devices and digital infrastructures. These real-time data can be utilised across various domains for numerous applications. These include, for instance, dynamic route optimisation and immediate intervention in the case of accidents; guidance for emergency vehicles and evacuation procedures during disasters; real-time repositioning of buses, taxis, and other transit options to adapt to dynamic travel demand fluctuations; real-time crowd management during special events; and urgent public health warnings, to name but a few. In this respect, Chapter 4 explored the potential of combining traffic data with complex network theory for resilience purposes. In particular, this synergy enables continuous surveillance of transportation networks, thus facilitating the quick identification of unexpected and difficult-to-predict disruptions [198, 213].

However, to unlock the full potential of both real-time data streams and extensive repositories of historical data, robust and heterogeneous technological support is imperative. From a software architectural point of view, it is necessary to provide flexible, scalable, and reliable solutions to accommodate the multifaceted needs that can arise in urban environments. Proper software infrastructure is required to manage a diverse array of envisioned services across various domains, which are often difficult to anticipate but usually rely on the same sources of data and share the same basic non-functional needs. Moreover, this support framework must provide general-purpose and complex data modelling capabilities to provide the necessary building blocks to implement the most variegated city applications from the available data. Importantly, it should be adept at managing the massive volumes of data generated within urban settings, especially in real-time scenarios. Finally, the infrastructure should be capable of adapting to fluctuating levels of usage load for these applications, ensuring that services remain accessible and responsive at all times.

Such support could thus take the form of a platform designed for the continuous and

seamless life-cycle of services, from design and development to deployment and operation. Over the past decade, several platforms designed for specific problems in various application domains have been introduced. Despite some advancements, we identified a clear gap[1] for general-purpose city-oriented platforms that could be highly adaptable and easy to customise for supporting city resilience according to a complete data-driven perspective.

Starting from these needs, in this chapter, we resume the research on software architectural, algorithmic and technological solutions carried out within the ANR project PROMENADE, culminated with the development of an open-source prototype platform of the same name[2]. The platform was designed to facilitate the monitoring and management of large-scale urban infrastructures towards data-driven resilience and sustainability, focusing on a transportation-related prototypical implementation. In addition, we developed and integrated into the platform two solutions for real-time monitoring and management of large-scale traffic networks, *i.e.,* rapid computation of betweenness centrality and real-time traffic zonal control. These solutions, based on the research results described in Chapter 4, represented two case-study applications of the proposed platform that confirmed its relevance for resilience enhancement and allowed for its performance evaluation.

The chapter is organised as follows. Sec. 5.1 details the requirements and modelling approach adopted for the architectural design of the PROMENADE platform. The platform global architecture and its main components are detailed in Sec. 5.2. The implementation of the platform core and its specialisation for large-scale road traffic networks is described in Sec. 5.3. This section also describes the realistic case studies that we considered for the evaluation of the platform prototype, by relying on road traffic datasets in the context of real-time road network monitoring. Sec. 5.4 describes our contribution to the exact computation of the betweenness centrality metric on very large-scale networks, a metric that we used for road monitoring purposes. Finally, Sec. 5.5 describes the proposed multi-agent cooperative traffic zonal control solution.

The results presented in this chapter have been the outcome of an intense research collaboration with University of Sannio, Italy, and have been at the core of the work of two PhD students (Cecile Daniel and Lorenzo Goglia) and two postdoctoral fellows (Mohammed Amine Merzoug and Pierre Lemaire) whom I have co-advised.

This chapter includes content from the following papers:

1. C. Colarusso, A. De Iasio, A. Furno, L. Goglia, M. A. Merzoug, and E. Zimeo, "PROMENADE: A big data platform for handling city complex networks with dynamic graphs", Future Generation Computer Systems, vol. 137, pp. 129–145, 2022.

2. C. Daniel, A. Furno, L. Goglia, and E. Zimeo, "Fast cluster-based computation of exact betweenness centrality in large graphs," Journal of Big Data, vol. 8, no. 1, pp. 1–39, 2021.

3. C. Daniel, P. Lemaire, A. Ladino, A. Furno, N.-E. E. Faouzi, and S. Hassas, "COMFORT: Cooperative multi-agent framework for large-scale routing-based traffic control", in 102nd Annual Meeting of the Transportation Research Board (TRB), 2023.

5.1 Modelling Approach and Platform Requirements

Cities are highly complex environments with diverse technical infrastructures co-existing and interacting continuously. For instance, city lights and water infrastructures are essential for the proper functioning of the road system. In addition to this structural interdependency, monitoring the vulnerabilities of a given city also means considering its inherent dynamism:

[1]The interested reader can refer to the related work section from [214].
[2]The source code of is available on GitHub, https://github.com/licit-lab/promenade_platform

an electric blackout should be promptly signalled to warn vehicles entering the affected area; the same may be said for emergencies due to floods.

According to the circumstances, city requirements may: *(i)* frequently change on a temporal basis, and *(ii)* be relaxed or tightened to meet changing Service Level Agreements (SLAs). In technical terms, approaching the complex needs of a city means fast supporting the elicitation of new service requirements and satisfying the ever-changing non-functional ones. This section discusses these requirement classes and other helpful general features a city platform should be equipped with.

Moreover, city complexity requires factorising the infrastructures and related behaviours in a unifying meta-model to reduce the platform complexity by focusing on a specific set of functional and architectural choices and quickly extending it to address new and changing needs. By harnessing the power of abstraction, these tasks can be eased by adopting an underlying unified meta-model. The considerations reported in Chapter 4, motivated us to adopt a graph-based meta-model, as will be further detailed below.

5.1.1 Complex network modelling of city infrastructures

Natural and human-engineered systems can be represented as a set of *complex time-changing* entities that interact with one another through *multiple complex time-changing* relationships (both entities and relationships can appear, disappear, and change their attributes over time), that originate *complex networks*.

We consider as complex network a graph that has numerous non-trivial structural features (*e.g.,* multiple attributes on edges and nodes, multiple edges between the same pair of nodes, attributes that change value over time, *etc.*) and can represent different real-world systems such as city, social, biological, computer, and technological networks [215, 216]. Moreover, complex networks can also include other types of features such as *(i)* multiple layers of semantically-different interactions between the modelled entities, and *(ii)* multiple intra-layer sub-networks, each corresponding to a group of entities with specific properties (*e.g.,* entities within a delimited geographical area, entities belonging to a specific organisation, *etc.*).

PROMENADE has been designed to perceive city networks as *dynamic complex networks* in which *nodes* (vertices) represent different homogeneous or heterogeneous city entities, and *links* (edges) correspond to the diverse relationships existing between these entities. Although a platform centred around the concept of complex network can be exploited to monitor and study numerous *complex dynamic systems* in disparate contexts, we mostly focused on urban technical networks (and specifically transportation), which can be organised in three levels of abstraction (see Fig. 5.1):

- At level 0, the bare physical space, which is represented by different (interdependent) city complex systems, like, for instance, transportation infrastructures (road networks, buses, trams, and metros), telecommunication/phone networks, electric grids, person networks, *etc.*
- At level 1, the deployed IoT, Edge, Fog, and Cloud nodes, which are responsible for providing the necessary technological resources to collect, transfer, store, and process data coming from multiple real-time sources integrated with the complex systems from level 0, at high-frequency, fine-resolution, and large spatio-temporal scales [217, 218].
- At level 2, complex network models and the different application services for knowledge extraction, data visualisation, machine learning, *etc.*

At the topmost level (Fig. 5.1), the city-system under analysis is abstracted as a complex multi-modal, multi-layer attributed time-varying network where nodes can have a hierarchical organisation (forming multiple physical or logical layers). These different layers can be utilised to model specific perceptions of the same system or a combination of different systems. For example, if we consider transportation systems, the obtained network can be

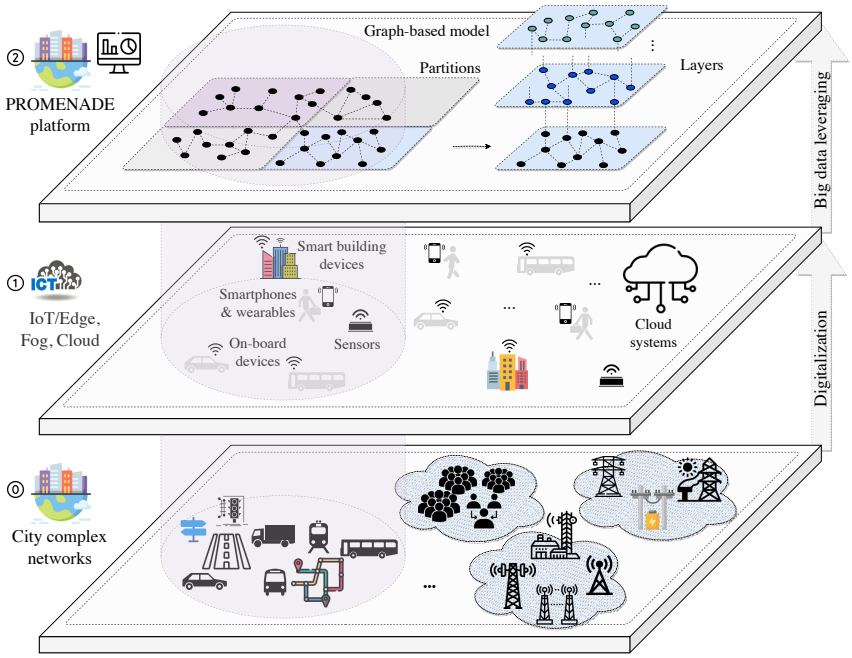


Figure 5.1: Collecting and leveraging big smart city data as complex dynamic networks.

a multi-layer combination of road and urban transit networks, including buses, trams, and metros. The considered city-system (each of its layers) can also be divided into numerous geographical *areas* or logical *groups* that identify specific sets of nodes and links (Fig. 5.1: level 2).

In addition, both nodes and links are characterised by multiple attributes belonging to two types: *(i) static* attributes representing the monitored system (*e.g.,* road network topology and its basic features), and *(ii) dynamic* attributes obtained through different continuous operations such as data sampling, aggregation, and computation (*e.g.,* average travel speed observations).

- *Static attributes*: to ensure the proper functioning of the platform, nodes and links must have a non-empty set of static descriptive characteristics that depend on the considered application. Along with these attributes, both nodes and links can have any additional number of attributes, depending on the entities they represent. For instance, nodes and links must have a unique identity (*nodeID* and *linkID*). As a second example, when entities of level 0 (see Fig. 5.1) are non-mobile (have a fixed location) in the physical space, their representative nodes can be characterised by two or three static parameters, depending on whether a 2D or 3D space is considered: *longitude*, *latitude*, *altitude*. In some cases (*e.g.,* road networks), links can also be located using the nodes that delimit them: *fromNode*, *toNode* (which are the *ID*s of the *adjacent* nodes of each link[3]).
- *Dynamic attributes*: each operation performed on nodes and links (*e.g.,* data sampling, graph metric computation, *etc.*) to measure a dynamic attribute must include at least two elements: *(i)* the *identity* of the entity to which the operation is associated and *(ii)* the *timestamp* at which this operation was executed. This means that nodes and links must be uniquely referenced through an identifier (*ID*). An example of attributes falling into this category is the dynamic location of mobile entities of level 0.

[3]We recall that any two nodes connected by an edge or any two edges connected by a node are said to be adjacent.

5.1.2 Functional requirements

The graph-based meta-model can be exploited to implement the diverse inter-connected systems composing a city in its entirety. By running specific standard or custom graph algorithms, the platform enables to *(i)* analyse a city partially or as a whole, *(ii)* analyse a single subsystem or a combination of them, and *(iii)* rapidly tackle emerging problems by setting properties on the graphs.

According to this main pillar, we identified a (minimal) set of functional requirements a graph-based smart city platform should satisfy:

- *Static data upload and update*: the platform must be able to upload and update the essential static data (initial graph structure), which represent the main entities of the targeted city system and their relationships.
- *Dynamic data acquisition and storage*: the platform should be able to continuously ingest node and link dynamic attributes from different sources. The collected dynamic data must be efficiently stored to be easily accessed via multiple frequent queries.
- *(Graph-based) data processing support*: the gathered big data may be useless if not transformed into meaningful and valuable information. Therefore, the platform must provide the necessary general-purpose (stream and batch) processing solutions to close the semantic gap between raw data collected from the environment and the information needed by the targeted services. At the same time, the processing middleware should be flexible enough to easily integrate additional processing services. Moreover, the platform must also provide graph-based data processing support. This set of functions or graph algorithms (such as depth-first search, breadth-first search, Dijkstra single-source shortest path, betweenness centrality, degree centrality, *etc.*) can be offered by specific graph-based software or implemented using the general-purpose processing engine.
- *Data visualisation*: the platform must be able to provide a user-friendly, map-based Graphical User Interface (GUI) to visualise city infrastructures, their digital representations, as well as the collected dynamic data, and to enable interaction with the platform itself.

5.1.3 Non-functional requirements

The adoption of a graph meta-model for a large-scale system requires satisfying some non-functional requirements as the model grows in complexity, such as:

- *Time efficiency*: the objective is to offer smooth, timely online and offline services to final users. To do so, in addition to the possible pre-processing of Edge and Fog nodes, the different implemented cloud software components must be highly time-efficient in terms of data ingestion, storage, processing, and querying (both live-streaming and historical);
- *Availability*: the proposed architecture with all its components and services must be fault-tolerant. That is, regardless of software and hardware failures, the system must be able to continue delivering its requested functionalities, and recovery must be instantaneous and transparent (without data loss or service interruption);
- *Scalability*: the proposed components must be highly scalable so that they can maintain the required low-latency performance in front of both an increasing data volume (large-size or large-amount of data attributes associated with graph entities) and increasing velocity (high-throughput data);
- *Flexibility*: the system should be easily changed by adding, updating, or removing software components at any time without interrupting the currently provided services. This point is fundamental as one cannot foresee and provide in advance all possible services that can be useful for a city.

By pairing DevOps, a methodology as well as a set of tools and practices that *merge* and *automate* development and IT operations, with the adoption of the cloud-native microservices architectural style [219, 220], the effort to implement these requirements can be

lessened.

5.1.4 Additional features

Other requirements depend on the applications and can be satisfied by orthogonal layers. In general, a plug-in mechanism should be provided to decouple the platform core from every end-user service that may be needed. The following are examples of possible additional features for some applications.

- *Data privacy*: the collected IoT data can be personal and sensitive. They cannot be utilised in their arrival raw state and must be pre-processed before use (anonymised, filtered, *etc.*). Across Europe, as already mentioned in Sec. 3.4, to solve this issue and harmonise data protection and privacy, the EU has specified a set of strict regulations and rules, called GDPR[4], which all digital businesses and services (using personal sensitive data) must comply with. To anonymise and filter the received data according to the consent of customers, the latter can be offered an online contract that lists all their preferences (*e.g.,* who can access their data, for how long, *etc.*), and allows them to access and modify these preferences at any time. The solutions proposed in Sec. 3.4 based on $k-$anonymity are an example in this direction.
- *Security*: whether in the case of collecting personal-sensitive information (such as personal identification and localisation) or not, security mechanisms (*i.e.,* confidentiality, integrity, and availability of information) should be provided.
- *Transaction integrity*: some applications, such as the ones that deal with citizens data for city services fruition, should ensure the collection of data as shared and immutable transactions in order to avoid repudiation of the originated data.

Even though we acknowledge the importance of the aforementioned additional requirements, these features and those not mentioned here were not specifically addressed in the prototypical implementation of the platform, being left for future work and platform extensions.

5.2 PROMENADE architecture

PROMENADE is a set of interconnected microservices that can be operationally divided into four *layers* and one *plane* (Fig. 5.2). The platform Initialization and Monitoring Plane (IMP) is a set of software components dedicated to initialise and monitor the platform and its different layers. It is responsible for uploading and updating the main static data of the monitored system. It is also responsible for monitoring the health of the cloud nodes, the hosted software components, their CPU, memory, and network usage, *etc.* All these features can be offered to end-users/administrators via a Web-GUI, through which they can initialise, monitor, and control the platform, its static and dynamic data, and its overall status.

The first three layers, which have been divided into stream and batch components, are in charge of big data ingestion, storage, and processing: *(i)* Stream data Ingestion (SI) and Historical data Ingestion (HI), *(ii)* Distributed Asynchronous producer-consumer Queue (DAQ) and Permanent Storage (PS), *(iii)* Stream Processing (SP) and Batch Processing (BP). The fourth Service offering layer (SVC) can query either or both the stream and permanent storage (DAQ and PS), which can contain raw and processed data, to provide the desired services to end-users. According to the Lambda architecture [221], PROMENADE handles both live streams and historical data. However, due to the nature of the applications, it mainly works on time windows of different sizes, often relying on Kappa architecture principles [222] at the implementation level.

[4]General Data Protection Regulation; applicable as of May 25, 2018, in all member states of EU (https://gdpr-info.eu/).

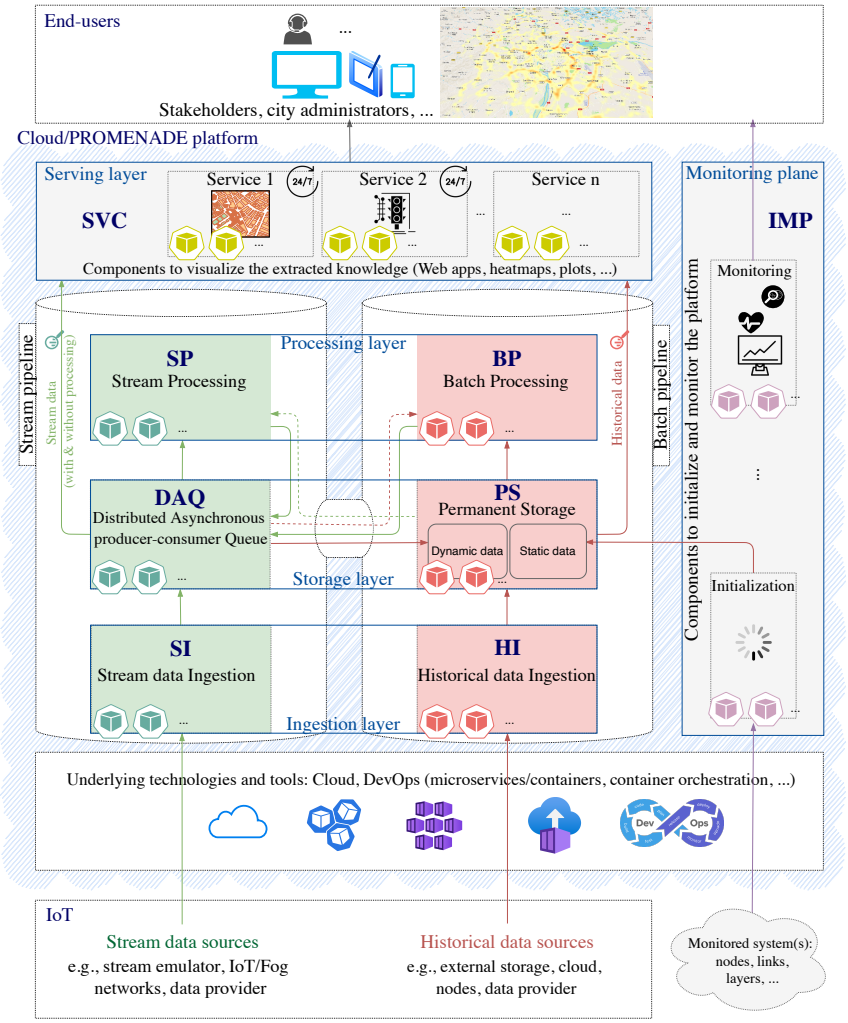


Figure 5.2: Detailed architecture of PROMENADE.

In the following, we detail the layers of PROMENADE in this order: (1) SI and HI, (2) DAQ and PS, (3) SP and BP, and finally (4) SVC.

5.2.1 Live-stream and historical-data ingestion

Once the main software components (*e.g.*, microservices) have been deployed and the platform properly initialised with the underlying data (static attributes of nodes and links), it is ready to start ingesting dynamic data and performing its required tasks. To achieve this goal, the main role of this ingestion (SI and HI) layer is *(i)* gathering (stream and historical) data from the defined sources (using predefined or custom software components) and *(ii)* preparing this input by making it usable by the upper layers. The HI is a specific software component responsible for ingesting pre-recorded dynamic data from external historical sources, while the SI provides live stream data ingestion from IoT devices. For instance, if the adopted communication protocol is MQTT[5], the SI component can be a high-performance MQTT broker responsible for receiving, decoding, formatting, and handing the output to the upper DAQ component [223].

[5]MQTT: The Standard for IoT Messaging, https://mqtt.org

5.2.2 Temporary and permanent storage

Raw and processed data are stored in two modes: *temporary stream queuing* and *permanent historical storage*.

- *Temporary stream queuing*: implemented through a DAQ, this component can mediate the interaction between every other platform component, thus enabling an asynchronous communication model. For example, the DAQ can receive both raw data coming from the bottom ingestion (SI) components and processed data generated by the top processing (SP and BP) components, making them consumable by any other interested component.
- *Permanent storage*: the PS component can be provisioned from three data sources: *(i)* initially by the IMP plane (with the structure and characteristics of the monitored system, *i.e.,* static node and link attributes), *(ii)* continuously from the previous DAQ component located to its left in Fig. 5.2 (dynamic data continually flows from the temporary space to the dedicated permanent storage), or *(iii)* at any time from the historical data ingestion (HI) component located below it. For the received *static* and *dynamic* data to be efficiently accessed and analysed at any subsequent time, we designed them to be separately stored in two distinct non-ephemeral PS components (Fig. 5.2), and therefore undergo different operations. Concerning static data (*i.e.,* the nodes and the links, as well as their static attributes), they can be created, updated, and deleted. This is not the case for dynamic data (*i.e.,* collected or computed attributes); they are recorded as immutable append-only.

5.2.3 Stream and batch processing

Data are sent to the platform to be processed for immediate (stream) or subsequent (historical) use. These two processing modes are based on the previously presented temporary and permanent storage (Sec. 5.2.2).

- *Stream processing*: the SP is responsible for ensuring low-latency, high-throughput retrieval and processing of continuous IoT data streams temporarily stored in the lower DAQ component. The SP is also responsible for data pre-processing. According to the considered application, the continuously ingested streams can be sent to this component to undergo all necessary beforehand cleansing, filtering, or aggregation. This stream processing component can retrieve data from both the stream and permanent (DAQ and PS) storage, but can only write (the obtained results) to the stream queue (DAQ). For instance, after pre-processing stream data, the modified data can be returned to the DAQ, and from there stored in the dedicated permanent storage (Fig. 5.2). This architectural choice provides *time* and *synchronisation decoupling*, which among other things help prevent backpressure issues.
- *Batch processing*: the BP is responsible for processing historical data. Similar to SP, this component can also read data from both the permanent and temporary storage (DAQ and PS), and can only write to the stream DAQ. As examples of possible tasks assigned to this component, we mention searching, querying, and applying different machine learning algorithms (classification, clustering, anomaly detection, *etc.*) to the large available historical data.

5.2.4 Service offering

Services can perform data *processing* or data *reporting*, where the latter may depend on the former. In particular, *processing* services, deployed at the level of the stream and historical pipelines, further process the incoming data and either expose their insight to the upper layer or persist them back into the storage components. On the other side, *reporting* services, which are either consuming raw data or exploiting the *processing* services output, offer fully-fledged dashboards.

5.3 Platform implementation and its specialisation in road network monitoring

This section discusses the implementation of the platform core framework, called PROMENADE-v2.0 (Sec. 5.3.1), and then tackles the specialisation of this core for road network monitoring (Sec. 5.3.2). The empirical evaluation of the platform performance is not reported in this document for brevity: readers can refer to the full paper [214] for a detailed presentation of the conducted experiments. This study allowed us to assess the scalability of PROMENADE-v2.0, and to set the main architectural parameters that significantly impact the performances of the proposed solution.

5.3.1 Platform core implementation

PROMENADE-v2.0 is the evolution of two previous versions proposed in [224] and [225]. In the following, we first introduce the underlying software and hardware technologies on top of which this core has been implemented and then present its main components and their technological choices. It is worth underlining that the development of the platform is an ongoing process and that only a part of the components has been fully implemented in the prototype, as graphically detailed in Fig. 5.4.

Underlying software and hardware technologies

The deployment environment of PROMENADE-v2.0 is a software/hardware infrastructure composed of:

- *An open-source middleware* built around *Docker*[6], *Kubernetes*, and *Red Hat OpenShift* (Fig. 5.3), the currently most popular DevOps tools to create and orchestrate containers.

[6]Docker: `https://www.docker.com/`

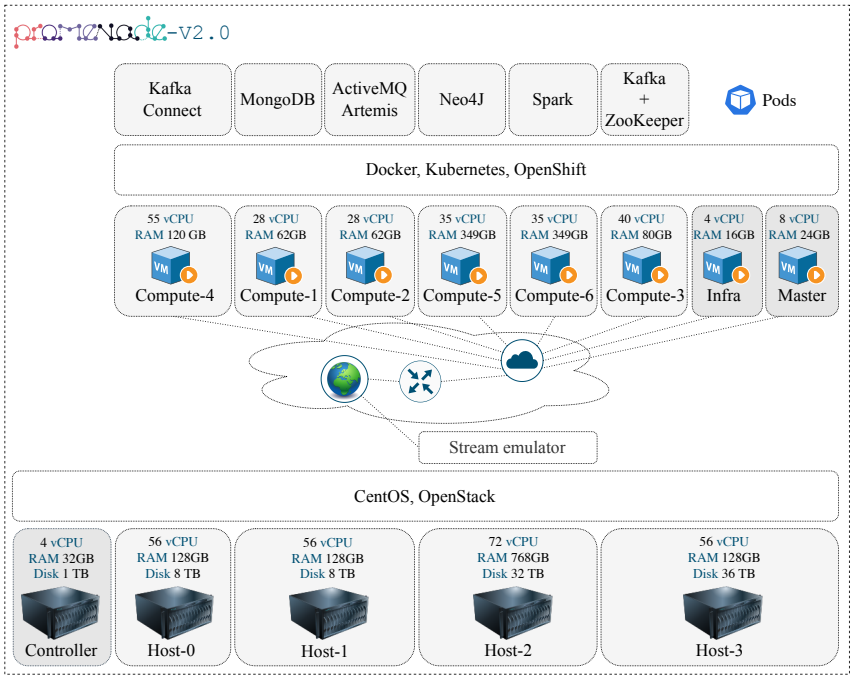


Figure 5.3: Underlying software and hardware technologies of PROMENADE-v2.0.

- *A virtualization environment* built atop OpenStack[7] that provides services for provisioning virtual machines (Nova), providing networking capabilities (Neutron) and managing disks (Cinder) to offer a block storage system.
- *Five physical machines* to host the OpenStack services and virtual machines. One of the physical machines operates as an OpenStack *controller* (with 1 CPU at 3.20GHz with 4 cores/4 threads), while the other four are configured to work as either *Compute* or *BlockStorage* nodes (each with 2 CPUs Intel Xeon, running at 2.20 or 2.40 GHz, more than a hundred of GB of RAM, TBs of hard disk, and two 1 Gb/s Ethernet adapters). The characteristics of the *controller*, *physical hosts* (Host 0, 1, 2, and 3) as well as their corresponding *guest VMs* (Master, Infra, Compute 1, 2, 3, 4, 5, and 6) are summarized in Fig. 5.3.

The overall infrastructure provides dynamically provisioned computing units, called *Pod*s, that can contain one or multiple *containers*. To satisfy the desired non-functional requirements, the infrastructure is specialised to govern the way such Pods are built, deployed, and kept in their desired state. This is performed by different layers of controllers and mechanisms such as: *ReplicaSet* (self-healing, scalability, ...), *Deployment* (updates and rollbacks), *StatefulSet*, *Jobs*, *Services*, *Routes*, *Custom Resource Definition* (CRD), *Operators*, *etc.* We redirect the reader to the specific technical documentation to have more details about the concepts introduced above.

The main role played by each component of PROMENADE has been discussed in Sections 5.2.1 to 5.2.4. In the following, we discuss the implementation of these components using open-source software technologies, which have been chosen not only because of their popularity, stability, and efficiency, but also because they comply with the design objective of PROMENADE. Tab. 5.1 compiles and maps the components of PROMENADE-v2.0 to their corresponding Kubernetes objects.

Live-stream ingestion

As Fig. 5.4 shows, Apache ActiveMQ Artemis[8] is the stream ingestion tool used to collect dynamic data from the lower IoT/Edge networks. To face any increase/decrease in the size of the monitored network, the deployed Artemis broker instances can be scaled up and down respectively (manually or automatically as needed). Also, to maintain the persistence and consistency of the data in case of possible crashes of the broker instances, a *StatefulSet* controller has been used (Tab. 5.1). These two points apply to any other component of the architecture that needs such control.

[7] OpenStack: https://www.openstack.org/
[8] Artemis: https://activemq.apache.org/components/artemis/

Component	K8s object
ActiveMQ Artemis	Stateful Set (STS) with 2 Persistent Volume Claims (PVCs) of 8 GiB each.
Neo4J	Three read-write cores (STS) and one read replica (Deployment). The three read-write cores consist of 1 leader and 2 followers with 3 PVCs of 10 GiB each.
Kafka	STS with 3 brokers/pods (0, 1, and 2) that have each one PVC of 10 GiB.
ZooKeeper	STS (0, 1, and 2) with 6 PVCs of 10 GiB each.
Kafka Center	Deployment.
kafka-Artemis source connector	Deployment.
Kafka-Mongodb sink ingestion/processing connector	Deployment.
MongoDB	Deployment with 1 PVC of 20 GiB.
Spark	The considered cluster is composed of one master. (CPU: 1 core, RAM: 1 GiB) and one worker (CPU: 32 cores, RAM: 200 GiB).

Table 5.1: Mapping of PROMENADE-v2.0 components to their corresponding K8s objects.

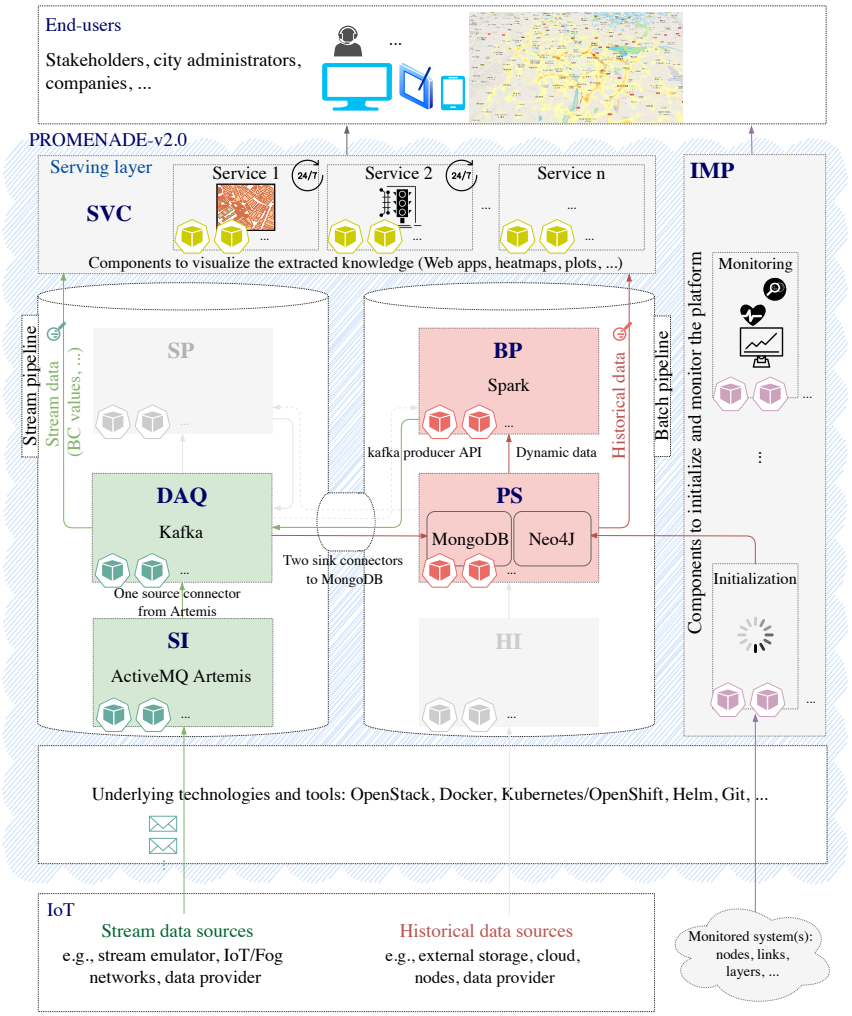


Figure 5.4: Detailed architecture of PROMENADE-v2.0. Gray boxes indicate that the corresponding components are not implemented in the current prototype of the platform.

As discussed in Sec. 5.1.1, the considered city-system (at each layer) can be split into logical groups (*e.g.,* geographical areas) of nodes and links. To implement this design, we considered a one-to-one mapping between the groups and the internal data structures of SI, DAQ, and PS components. For instance, a group corresponding to a geographical area in a city can be mapped to an Artemis queue (SI component). It is worth remembering that these groups can be defined via the initialisation component from the IMP plane. The implementation details about the one-to-one mapping of the DAQ and PS components will be presented in the next subsections. This partitioning configuration allowed us to achieve better performances.

Temporary queuing storage

The popular Apache Kafka[9] and its discovery service ZooKeeper have been considered to comply with the design of the queuing component. To deal with data persistence while providing fault-tolerance, a StatefulSet controller has been used for both Kafka and ZooKeeper, as is the case for Artemis (Tab. 5.1). This way, if one of the stream brokers goes down, it will be recreated anew while retaining the old session.

[9]Kafka: https://kafka.apache.org/

In accordance with the *temporary* nature of this component, data is not permanently stored: depending on the chosen retention policy, data can be retained for one hour, one day, one week, *etc.* However, an important point about data-flow concerns the *source* and *sink* connectors, which are respectively in charge of getting data in and out of Kafka. We used three connectors to implement PROMENADE-v2.0 (Fig. 5.4): one customized source connector (which links Kafka to Artemis) and two out-of-the-box sinks (which both connect Kafka to MongoDB for ingestion and processing, respectively). The role of MongoDB is explained in the next subsection.

The adopted Kafka source connector was implemented through an adapter that retrieves packets from Artemis queues, uniforms their format, and forwards them to Kafka topics. The two sink connectors work as follows: when Kafka receives data from the Artemis-based SI component, the first connector writes the received dynamic data of links and nodes (*e.g., average travel time* data of links) from Kafka to MongoDB, while the second is responsible for writing the computed node and link attributes (*e.g., betweenness centrality* of nodes or any other information computed at the SVC level) from Kafka to MongoDB.

As done for Artemis (SI component), to implement the logical grouping of city entities, also this DAQ component requires a specific topic to be associated to each group.

Permanent historical storage

As detailed in Sec. 5.2.2, PROMENADE-v2.0 handles static and dynamic data with two distinct non-ephemeral components. To implement this design, we opted for two popular technologies: Neo4J[10] and MongoDB[11], for static and dynamic data, respectively.

The former is an open-source NoSQL graph DBMS implemented in Java that supports efficient graph-based queries of static data (structure and characteristics of the monitored system). To this end, Neo4J uses Cypher, a declarative query language similar to SQL but optimized for graph databases.

MongoDB is an open-source NoSQL document DBMS, developed in C++ around two main concepts: *documents* and *collections*. A *collection* is a group of documents, and a *document* is a data structure composed of field-value pairs (JSON objects). We used MongoDB for dynamic data because samples in the form of JSON documents can be immediately stored, while embedded documents and arrays reduce the need for expensive joins. Moreover, the dynamic schema support, which means that collections do not enforce any schema, ensures high flexibility for handling different kinds of data.

For each logical group of city entities, we created specific MongoDB collections. Each of these collections is ultimately in charge of storing the node and link samples coming from the stream ingestion component, and the node and link attributes computed by the data processing component.

Data processing

To satisfy the functional requirements listed in Sec. 5.1.2, PROMENADE-v2.0 offers both *general-purpose* and *graph-based* data processing support. The former has been implemented by leveraging a general-purpose engine for large-scale data processing. Specifically, as shown in Fig. 5.4, we opted for the well-known Apache Spark[12] framework, which can work on both data streams and data at rest. This way, by injecting appropriate (general-purpose or graph-based) algorithms into this data processing component, the platform can be easily extended to provide end-users with fine-grained services catering to the different problems that should be solved.

[10]Neo4J: https://neo4j.com/
[11]MongoDB: https://www.mongodb.com/
[12]Spark: https://spark.apache.org/

The basic functions of the graph-based support are provided by Neo4J (which is used for storing and querying static data). For instance, the library of Neo4J contains implementations of path discovery and centrality algorithms[13]. The first category helps to identify shortest paths and perform search tasks (*e.g.,* depth-first search, breadth-first search, Dijkstra single-source shortest path, *etc.*), whereas the second determines the importance of distinct nodes in a network (*e.g.,* betweenness, degree, eigenvector centrality, *etc.*). Other examples of provided algorithms include community detection, similarity functions, topological link prediction, node embeddings, node classification, and link prediction.

5.3.2 Specialisation for Road Network Resilience

To assess the flexibility of the proposed platform architecture and prototype, we report in this section how PROMENADE-v2.0 has been specialised for real-time road traffic monitoring and management applications.

Data model

Instantiating the general platform means *(i)* instantiating the graph-based meta-model (of Sec. 5.1.1) into a specific model related to the considered specialisation, and *(ii)* providing the necessary dynamic (node and link) values for this graph through appropriate data sources, whereas the desired non-functional requirements are satisfied as a consequence of the application-level choices and the available hardware resources. In particular, we specialised PROMENADE-v2.0 for *road traffic monitoring and management* by *(i)* modelling the monitored road infrastructure as a *one-layer directed multi-attributed* graph, and *(ii)* consuming the constantly changing traffic conditions (dynamic observations) retrieved from its links.

Formally, the road network graph is denoted as $G_{RN} = (V, E, \mathcal{A}, \mathcal{B})$, where V is the set of vertices or road intersections, E is the set of edges or road segments, $\mathcal{A}$ is the set of attributes associated to nodes, *i.e.,* $\{\mathcal{A}_v | v \in V\}$ and $\mathcal{B}$ is the set of attributes associated to the edges, *i.e.,* $\{\mathcal{B}_e | e \in E\}$. The attributes associated to nodes and edges in such a graph could be static (not frequently changing) or dynamic. More specifically, as shown in Tab. 5.2, nodes static attributes include: a *unique identifier* (*osmID*[14]), a *latitude-longitude* value, and their *area name*. On the other hand, edges static attributes include: *linkID*, *fromNode*, *toNode*, *street name*, *street length*, *speed limit*, *estimated free-flow speed*, *area name*, and *coordinates*; *fromNode* and *toNode* are the *osmID* of the adjacent nodes of an edge and *coordinates* is its geometry. The choice of the attributes listed above is dictated by the available fields in the datasets used in our case study evaluation. However, nodes and links can have other static attributes depending on the specific application. For instance, our dataset allows further characterising a link via the *width* (in *meters*), the *number of lanes*, the presence of *bridge* or *tunnel* for the associated road segment. These attributes were ignored in this work as irrelevant to our case study.

As an additional feature, the single layer of our graph, *i.e.,* G_{RN}, is composed of multiple sub-networks G^z_{RN}, each representing a different geographical area $z \in \mathcal{Z}$ (from a given segmentation of the modelled city) identifiable via the *area name* parameter. More formally, G^z_{RN} is the sub-network induced by the set of nodes $V^z \subseteq V$ and edges $E^z \subseteq E$ that are geographically located in the given area z.

The mandatory and non-mandatory *static* characteristics of road networks (Tab. 5.2) are injected into the implemented platform during the initialisation phase. Once configured and ready, the platform can then be continuously fed with *dynamic* attributes (in our case, *average travel time* of links, as from Tab. 5.3). The values of the link attributes can be thus used to produce insightful vulnerability indicators such as the betweenness centrality

[13] Neo4j Graph Algorithms: https://neo4j.com/developer/graph-data-science/graph-algorithms/
[14] OpenStreetMap ID (https://www.openstreetmap.org/).

Node attrib.	Type	Description
osmID	long	OpenStreetMap ID[14].
longitude	double	Longitude (x-coordinate in decimal degrees).
latitude	double	Latitude (y-coordinate in decimal degrees).
area name	String	Node area name.
Link attrib.	**Type**	**Description**
linkID	long	Link ID.
fromNode	long	OsmID of first adjacent node u.
toNode	long	OsmID of second adjacent node v.
name	String	Link (street) name.
length	double	Link length (*meters*).
speedLimit	int	Maximum speed (km/h); 0 if not applicable.
ffs	double	Estimated free-flow speed (km/h).
areaName	String	Link area name.
coordinates	List(double, double)	Sequence of (*lat, long*) points representing a street between two intersections (going from u to v).

Table 5.2: Road segments (links) and intersections (nodes): static data provided to PROMENADE-v2.0.

Link attrib.	Type	Description
linkID	long	Link ID.
timestamp	dd/mm/yyyy hh:mm:ss	Data aggregation time.
average tt	double	Average travel time (s).
Node attrib.	**Type**	**Description**
osmID	long	Node ID.
timestamp	dd/mm/yyyy hh:mm:ss	Data computation time.
BC	double	Betweenness centrality.

Table 5.3: Road segments (links) and intersections (nodes): dynamic data collected and produced by PROMENADE-v2.0.

of nodes (see Sec. 4.2.2), which represents, in turn, a time-varying attribute of nodes (as reported again in Tab. 5.3).

Stream data generation

The IoT networks correspond to the underlying level on top of which the platform was built (Fig. 5.4). Although being crucial, in the implemented version of our platform specialisation, they have been emulated by playing back an available offline dataset. This choice was motivated by two main reasons. First, the strong dependency of the IoT/Edge systems on the physical infrastructure of cities, and second, the flexibility of generating data as needed. The dataset used in the emulation to reproduce realistic road traffic conditions was the same as the one used in Sec. 4.2 for our experimental studies on vulnerability assessment related to the city of Lyon. This dataset was organised according to the structure presented in Tab. 5.2 and Tab. 5.3.

We implemented the emulation software visualised in Fig. 5.5: it can be *(i)* exploited and parameterised to stress and test the developed platform while considering extreme conditions (something that is not possible or very difficult in a physical environment), *(ii)* easily reused to study and analyse other complex networks for which offline historical data is available, and *(iii)* rapidly and straightforwardly substituted whenever physical IoT networks are available. Moreover, it does not reduce the overall validity of the platform. It is worth noting that the emulation focuses on data generation and does not consider edge-to-cloud communication delays, which were left out of scope for our prototype implementation.

From a technical point of view, the main task of the emulator is to consume the provided historical dynamic (raw) data (described in Tab. 5.4) and behave similarly to a set of Edge nodes.

Specifically, we coded the emulator to continuously produce the aggregate (average)

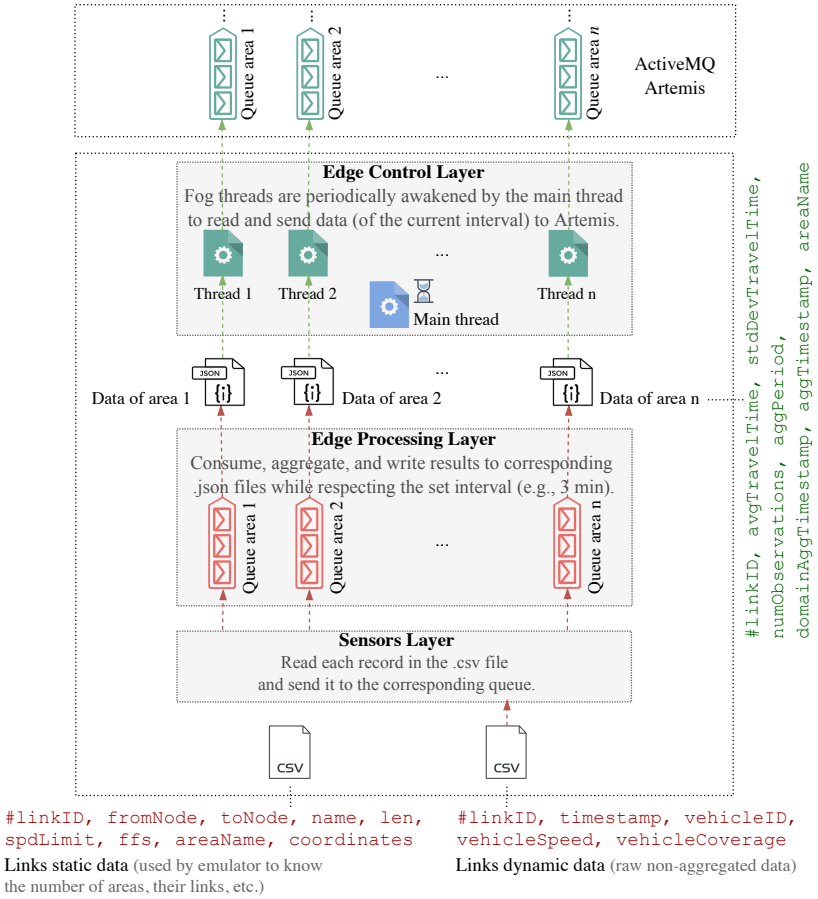


Figure 5.5: Architecture of the IoT data-driven emulation software.

travel time on links (Tab. 5.3) based on measures of speed observed from individual vehicles. To this end, the emulator was designed (see Fig. 5.5) to be a multi-threaded Java application that is composed of three main layers: *sensors layer*, *Edge processing layer*, and *Edge control layer*:

- *Sensors layer*: its role boils down to emulating data readings at the Edge. As shown in Fig. 5.5, it retrieves all samples (raw non-aggregated data of Tab. 5.4) and sends them as messages to the different queues of the Edge processing layer, based on the *area name* attribute from the link static data file of Tab. 5.2. In fact, the Edge processing layer hosts a queue for each geographic area. By this approach, we kept the geographic division inside the digital model by implementing it at the level of software components.
- *Edge processing layer*: the role of each Edge node of this layer is to consume raw samples (vehicle speed readings from sensors or on-board connected GPS devices) that are continuously coming to its respective queue (from the emulated sensors layer) and operates as follows. During the aggregation interval (*e.g.,* 3 minutes in our case), an emulated Edge node just handles a given vehicle sample by *(i)* increasing the number of observations for the link crossed by the vehicle (*link.numObservations++*), *(ii)* computing the vehicle travel time (*vtt*)[15], and *(iii)* adding the computed (*vtt*) value to the link statistics (*link.stats*). When the aggregation interval expires (after 3 minutes), the node computes and writes the aggregation result of its links (*e.g., avgTravelTime*, *stdDevTravelTime*, *aggTimestamp*, *etc.*) to a JSON file that refers to a specific area represented by the queue

[15] *vtt = (link.length * msg.vehicleCoverage) / msg.vehicleSpeed*

from which it consumes the raw data (Fig. 5.5).

- *Edge control layer*: mimics the behaviour of multiple *parallel* Edge nodes. More specifically, once the previous layer has finished its task of aggregating all data, this layer can start using its output JSON files. As shown in Fig. 5.5, this layer is composed of one *main thread* and n *Edge threads* (corresponding to the n areas of the city of Lyon). The role of the main thread is to control the online Edge threads by creating, starting, and synchronising them. Upon their creation, Edge threads sleep indefinitely and wait to be awakened. Once awaken, each Edge thread reads its corresponding aggregation file, sends data to Artemis, and then goes back to sleep (waiting for the next aggregation interval to come).

Link attribute	Type	Description
linkID	long	Link ID.
timestamp	dd/mm/yyyy hh:mm:ss	Data collection time.
v_ID	int	Vehicle or observation ID.
v_speed	double	Measured vehicle speed (km/h).
v_coverage	double	How much of the road segment has been covered by the vehicle when the sample was taken ($0 \leq coverage \leq 1$).

Table 5.4: Road segments: dynamic raw (non-aggregated) data of vehicles.

The emulation process has been designed to be configurable. It can be *realistic* (samples generated according to their original timestamps) or *virtual* (samples generated at a very high or low speed to enable several analysis scenarios). To be able to do so, a temporal factor is utilised to allow control of the frequency at which samples are generated while retaining their respective order. Moreover, each emulated Edge node can either compute the average travel time (of each link during a given time interval, *e.g.,* every 3 minutes), or it can behave differently depending on the considered application. For instance, it can act as a collector by forwarding raw samples to the Cloud platform.

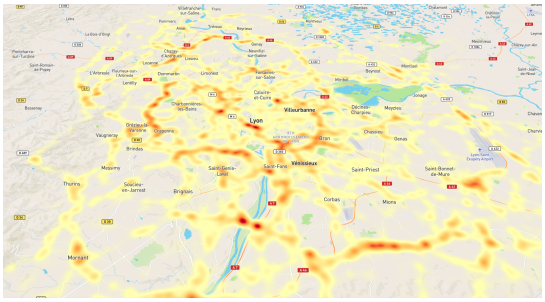
We recall that the one-to-one mapping between geographical/logical partitions and queues implemented in the emulated Edge processing layer is also adopted in the SI, DAQ, and PS components of PROMENADE-v2.0.

Implemented services

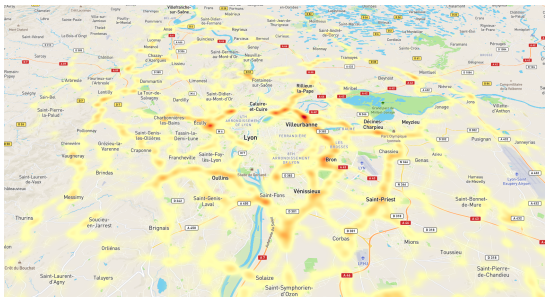
The specialisation of PROMENADE-v2.0 for road network monitoring provides two main classes of services: *real-time monitoring* and *traffic management*. We also implemented visualisation support for real-time presentation of the observed traffic conditions.

Real-time monitoring exploits the betweenness centrality as a metric to *continuously* detect critical elements of the network, *i.e.,* those characterised by higher values of centrality [213]. Introduced in [193], BC measures the importance of nodes or links for the flow of information in a graph. It is widely used to identify opinion leaders or influential people in social network analysis [226], vulnerabilities in computer networks [227], and threats from terrorist networks [228]. As discussed in Chapter 4, BC can also help detect critical intersections in transportation networks [213, 229, 230, 231, 232] and be used as a vulnerability indicator. The knowledge of critical nodes or links may enact appropriate control actions, provide relevant recommendations, and guide drivers through less vulnerable itineraries not traversing high-BC nodes.

Based on Brandes' algorithms [192, 233], the BC solutions provided by Neo4J are inefficient [213]. Thus, we used the general-purpose processing engine available in PROMENADE (Spark) to run our algorithmic approach proposed in [213] for approximate BC computation on directed and weighted graphs. The latter has been further enhanced in [234] for the exact computation of BC on undirected and unweighted graphs (see Sec. 5.4). These algorithms have been integrated as a core feature in PROMENADE-v2.0 to periodically compute the BC values of large-scale time-varying networks.



(a) Global state of the road network at 05:45:00 am.



(b) Global state of the road network at 05:45:00 pm.

Figure 5.6: GUI examples - Map with vulnerability indicator (BC). The hottest spots of the heatmap represent critical nodes with high BC values.

The collected dynamic data, together with the computed BC values (sent to Kafka and then stored in MongoDB), can be exploited by other additional data processing and visualisation services. We used two different connectors for collecting traffic conditions (*i.e.,* the average travel times) and for producing the real-time vulnerability indicators (*i.e.,* the BC values of nodes). The first connector is based on a RESTful API and operates in a pull mode, whereas the second exploits WebSockets and works in a push mode to provide continuous updates. The Kafka broker handles these operations: whenever an update/event occurs, it triggers the process of retrieving the necessary data and serving subscriptions. Fig. 5.6 reports an example of the output produced by the platform. BC values are represented in two different moments of a day using a heatmap layer, showing the evolution of the most central nodes over the day depending on varying traffic conditions.

The details of the adopted BC algorithm for real-time traffic monitoring are described in the next section, focusing on the solution proposed for rapid exact computation on large-scale undirected and unweighted networks [234], leaving the interested reader to our papers [235, 213] for a detailed description of the approximate versions for weighted and directed networks.

The solutions implemented for real-time traffic management are instead reported in Sec. 5.5.

5.4 Fast Cluster-based Computation of Betweenness Centrality in Large Graphs

Despite its great potential, BC requires a considerable computation time, which often represents a barrier to applying this metric in large-scale contexts, especially with dynamic graphs, whose topology or weights change frequently.

In recent years, the Floyd method [236], which requires $O(n^3)$ computation time, has

been overcome by the well-known Brandes' algorithm [192]. Given a graph $G(V, E)$, it exhibits $O(n + m)$ space complexity, $O(nm)$ time complexity for unweighted graphs, and $O(nm + n^2 log(n))$ for weighted ones, where $n = |V|$ is the number of nodes and $m = |E|$ the number of edges. However, the polynomial complexity of Brandes' algorithm is still an obstacle for analysing very large networks. Such a problem becomes even more evident and limiting if centrality is used for real-time (or quasi-real-time) analysis of dynamic networks.

By building on our previous work on approximated BC computation for unweighted-undirected graphs [237, 232] and weighted-directed ones [235, 213], we developed, as part of Cecile Daniel's PhD thesis, a novel algorithm (widely discussed in [238, 234]) which makes it possible to perform the exact calculation of BC on large, undirected graphs with a high speedup when compared to Brandes' algorithm and a significant improvement over recent variants of Brandes' algorithm based on clustering [239].

Undirected graphs are very common in real-world systems: examples include social networks, communication networks, protein interaction graphs, people interaction graphs, finite element meshes, and the topology of transport networks. Among these graphs, scale-free and Barabási-Albert graphs [240] represent an ideal candidate for the analysis of the proposed algorithm, since they model many real-world systems, such as the World Wide Web, the Internet and other computer networks, citation networks, social networks, airline networks, financial networks, *etc.* However, considering the application of our approach for real-time monitoring of road traffic networks, we also examined small-world networks by including in our benchmark the real-world road network of Lyon, France.

In the following, we introduce the notation used throughout the paper and briefly describe Brandes' algorithm. Then, we present our algorithm and the most relevant performance evaluation results. The presentation of the literature review, the details of the algorithm implementations (a Scala *map-reduce* parallel and a sequential one), the mathematical proofs and detailed performance analyses are left out of this document for brevity but can be found in the aforementioned publications [238, 234].

5.4.1 Background

Let $\mathbf{G}(\mathbf{V}, \mathbf{E})$ be an undirected unweighted graph with $\mathbf{V}$ representing the set of n vertices (or nodes) and $\mathbf{E}$ the set of m edges (or links). Let $s, t \in \mathbf{V}$ be two generic nodes of $\mathbf{G}$. We denote by $e_{s,t}$ the edge connecting s and t. The *neighbours* of a vertex s are all vertices u such that $e_{s,u} \in \mathbf{E}$. The *distance* between s and t, denoted by $d_{\mathbf{G}}(s, t)$, is the length of the shortest path(s) connecting them in $\mathbf{G}$. The *number* of shortest paths between s and t is denoted by $\sigma_{s,t}$, whereas the number of shortest paths between s and t that cross a generic node $v \in \mathbf{V}$ is denoted by $\sigma_{s,t}(v)$. It is worth noting that since the graph is undirected, $d_{\mathbf{G}}$ and σ are symmetric functions, thus $d_{\mathbf{G}}(s, t) = d_{\mathbf{G}}(t, s)$, $\sigma_{s,t} = \sigma_{t,s}$ and $\sigma_{s,t}(v) = \sigma_{t,s}(v)$. Given a generic node $w \in \mathbf{V}$, $\mathbf{P}_s(w) = \{u \in \mathbf{V} : e_{u,w} \in \mathbf{E}, d_{\mathbf{G}}(s, w) = d_{\mathbf{G}}(s, u) + 1\}$ is the set of direct *predecessors* of vertex w on shortest paths from s.

The *Betweenness Centrality* (BC) of a vertex $v \in \mathbf{V}$ is defined as follows:

$$BC(v) = \sum_{s \neq v \neq t \in \mathbf{V}} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}} \tag{5.1}$$

$BC(v)$ thus represents the fraction of shortest paths containing v among all the shortest paths in the graph between any generic pair of nodes s and t, summed over all possible pairs s and t with $s \neq v$, $s \neq t$ and $v \neq t$.

We refer to Tab. 5.5 for a summary of the notation used in the paper.

Notation	Description
$\mathbf{G}$	undirected unweighted input graph
$\hat{\mathbf{G}}$	a connected sub-graph of $\mathbf{G}$
$\mathbf{V}$	set of vertices of $\mathbf{G}$ ($\lvert\mathbf{V}\rvert = n$)
$\mathbf{V}_{\hat{\mathbf{G}}}$	set of vertices of $\mathbf{G}$ inducing $\hat{\mathbf{G}}$ (set of vertices of $\hat{\mathbf{G}}$)
$\overline{\mathbf{V}_{\hat{\mathbf{G}}}}$	set of vertices in $\mathbf{V} \setminus \mathbf{V}_{\hat{\mathbf{G}}}$
$\mathbf{V}_{HSN}$	set of vertices of HSN
$\mathbf{E}$	set of edges of $\mathbf{G}$ ($\lvert\mathbf{E}\rvert = m$)
$e_{s,t}$	edge connecting vertices s and t
$d_{\mathbf{G}}(s,t)$	distance between vertices s and t in $\mathbf{G}$
$\hat{d}_{\mathbf{G}}(s,t)$	normalised distance between vertices s and t in $\mathbf{G}$
$\sigma_{s,t}$	number of shortest paths between vertices s and t
$\sigma_{s,t}(v)$	number of shortest paths between vertices s and t which cross vertex v
$\hat{\sigma}_{s,t}$	normalised number of shortest paths between vertices s and t
$\mathbf{P}_s(v)$	set of direct predecessors of vertex v on shortest paths from vertex s
$\mathbf{P}_s(\mathbf{V})$	set of direct predecessors of vertices in $\mathbf{V}$ on shortest paths from vertex s
$BC(v)$	betweenness centrality of vertex v
$\delta_{s,t}(v)$	pair-dependency of the pair of vertices (s,t) on the intermediary vertex v
$\delta_{s,\bullet}(v)$	dependency score of vertex s on vertex v due to all destination vertices
$\delta_{s,\mathbf{V}_{\hat{\mathbf{G}}}}(v)$	dependency score of vertex s on vertex v due to all destination vertices in $\mathbf{V}_{\hat{\mathbf{G}}}$
$\mathbf{C}$	set of clusters of $\mathbf{G}$
$\mathbf{C}_i$	a generic cluster in $\mathbf{C}$
$\mathbf{C}(v)$	the cluster vertex v belongs to
$\mathbf{C}^*$	set of extended clusters in $\mathbf{G}$
$\mathbf{C}_i^*$	a generic extended cluster in $\mathbf{C}^*$
$\mathbf{K}$	set of all the equivalence classes
$\mathbf{K}_i$	an equivalence class
$\mathbf{K}_{\mathbf{C}_i}$	set of equivalence classes of cluster $\mathbf{C}_i$
$\mathbf{P}$	set of all the pivots
k_i	pivot node of the equivalence class $\mathbf{K}_i$
$\mathbf{EN}$	set of all the external nodes
$\mathbf{EN}_{\mathbf{C}_i}$	set of external nodes of cluster $\mathbf{C}_i$
$\mathbf{BN}$	set of all the border nodes
$\mathbf{BN}_{\mathbf{C}_i}$	set of border nodes of cluster $\mathbf{C}_i$
$\mathbf{BN}_{\mathbf{C}_i}(s,t)$	set of border nodes of cluster $\mathbf{C}_i$ on shortest paths from $s \in \mathbf{V}_{\mathbf{C}_i}$ to $t \in \overline{\mathbf{V}_{\mathbf{C}_i}}$
b_i	a generic border node in $\mathbf{BN}$
$\delta^{\gamma}_{s,\bullet}(v)$	global dependency score of s on v due to all $t \in \overline{\mathbf{V}_{\mathbf{C}(s)}}$ (same as $\delta^{\gamma}_{s,\overline{\mathbf{V}_{\mathbf{C}(s)}}}(v)$)
$\delta^{\gamma}_{s,\mathbf{V}_{\mathbf{C}(v)}}(v)$	global dependency score of s on v due to all $t \in (\overline{\mathbf{V}_{\mathbf{C}(s)}} \cap \mathbf{V}_{\mathbf{C}(v)})$
$\delta^{\gamma}(v)$	sum of all the global dependency scores (global BC) on v
$\delta^{\lambda}_{s,\bullet}(v)$	local dependency score of s on v due to all $t \in \mathbf{V}_{\mathbf{C}(s)} = \mathbf{V}_{\mathbf{C}(v)}$
$\delta^{\lambda}(v)$	sum of all the local dependency scores (local BC) on v
$\delta^{\epsilon}_{s,\bullet}(v)$	dependency score of s on v, as external node, due to all $t \in \mathbf{V}_{\mathbf{C}(s)}$
$\delta^{\epsilon}(v)$	sum of all the dependency scores on v as external node

Table 5.5: The notation used with the BC exact computation algorithm.

5.4.2 Brandes' algorithm

Brandes' algorithm is the fastest known general-purpose sequential algorithm for computing BC. It is based on the notions of *pair-dependency* and *dependency score*. Let us consider two generic nodes $s, t \in \mathbf{V}$. Given shortest paths counts $\sigma_{s,t}(v)$ and $\sigma_{s,t}$, the pair-dependency $\delta_{s,t}(v)$ of a pair s, t on an intermediary node $v \in \mathbf{V}$ is defined as follows:

$$\delta_{s,t}(v) = \frac{\sigma_{s,t}(v)}{\sigma_{s,t}} \tag{5.2}$$

The pair-dependency represents the fraction of shortest paths between s and t crossing v. The dependency score $\delta_{s,\bullet}(v)$ of a vertex s on a vertex $v \in \mathbf{V}$ is then defined as follows:

$$\delta_{s,\bullet}(v) = \sum_{t \in \mathbf{V} \setminus \{s,v\}} \delta_{s,t}(v) \tag{5.3}$$

BC can thus be redefined in terms of dependency score:

$$BC(v) = \sum_{s \neq v \neq t \in \mathbf{V}} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}} = \sum_{s \neq v \neq t \in \mathbf{V}} \delta_{s,t}(v) = \sum_{s \in \mathbf{V}} \delta_{s,\bullet}(v) \tag{5.4}$$

The key observation of Brandes' algorithm is that the dependency score obeys a recursive formula that considers the dependency scores of the successors of node v, *i.e.*, $\delta_{s,\bullet}(w)$. In particular, for each $s \in \mathbf{V}$ we have:

$$\delta_{s,\bullet}(v) = \sum_{w : v \in \mathbf{P}_s(w)} \frac{\sigma_{s,v}}{\sigma_{s,w}} \cdot (1 + \delta_{s,\bullet}(w)) \tag{5.5}$$

Brandes' algorithm runs in two phases, exploiting equation 5.5. For each (source) node $s \in \mathbf{V}$, in the first phase, a Single-Source Shortest-Paths (SSSP) algorithm, based on Breadth-First Search (BFS), is executed on $\mathbf{G}$ to find all the shortest paths rooted in s. In the second phase, dependency scores are accumulated by backtracking along the discovered shortest paths using the recursive relation in Eq. 5.5. In backtracking, nodes are visited in descending order of distance from the source. During these two phases, for each node $v \in \mathbf{V}$, the algorithm builds and exploits the following data structures: the set of direct predecessors $\mathbf{P}_s(v)$ on shortest paths from the source, the distance $d_{\mathbf{G}}(s,v)$ from the source, the number of shortest paths $\sigma_{s,v}$ from the source and the dependency score $\delta_{s,\bullet}(v)$ that accumulates the contribution of the source on node v due to all destinations during the back-propagation step.

5.4.3 Equivalence class

To reduce the number of SSSP explorations in the Brandes' algorithm, and thus lower the BC computation time, we exploited the concept of *equivalence class*. The algorithm leverages structural properties of graphs to find classes of equivalent nodes: by selecting one representative node for each class, it becomes possible to calculate BC by significantly reducing the number of SSSP explorations and thus accelerating its computation.

Formally, given a connected sub-graph $\hat{\mathbf{G}}$ of $\mathbf{G}$ induced by the set of nodes $\mathbf{V}_{\hat{\mathbf{G}}} \subset \mathbf{V}$, we define an equivalence class $\mathbf{K}_i$ as any subset of nodes in $\mathbf{V}_{\hat{\mathbf{G}}}$ that produce the same dependency score on all nodes - and for destinations - outside sub-graph $\hat{\mathbf{G}}$ when used as sources for SSSP explorations.

By choosing only one representative node (called pivot) for each class, the correct dependency scores of nodes can be computed by multiplying the scores computed via the SSSP rooted in the pivot by the cardinality of the class. Let k_i be a pivot of $\mathbf{K}_i$ and $v \notin \mathbf{V}_{\hat{\mathbf{G}}}$, *i.e.*, a node outside sub-graph $\hat{\mathbf{G}}$. According to our definition, we have:

$$\sum_{s \in \mathbf{K}_i} \sum_{t \notin \mathbf{V}_{\hat{\mathbf{G}}}} \delta_{s,t}(v) = |\mathbf{K}_i| \cdot \sum_{t \notin \mathbf{V}_{\hat{\mathbf{G}}}} \delta_{k_i,t}(v)$$

which, according to our notation, can be re-written as:

$$\sum_{s \in \mathbf{K}_i} \delta_{s,\overline{\mathbf{V}_{\hat{\mathbf{G}}}}}(v) = |\mathbf{K}_i| \cdot \delta_{k_i,\overline{\mathbf{V}_{\hat{\mathbf{G}}}}}(v) \tag{5.6}$$

with k_i representing the pivot node of equivalence class $\mathbf{K}_i$. Equation 5.6 clearly shows that a low number of classes significantly reduces the computation time, by allowing to skip a high number of SSSP explorations.

5.4.4 Clustering and BC computation

A possible technique to identify equivalence classes is to consider reference nodes. Given a generic sub-graph $\hat{\mathbf{G}}$, the reference nodes in $\mathbf{V}_{\hat{\mathbf{G}}}$ are those that need to be traversed to reach, via shortest paths from nodes in $\mathbf{V}_{\hat{\mathbf{G}}}$, any other node in $\overline{\mathbf{V}_{\hat{G}}}$.

To easily identify reference nodes, we used clustering, and to increase the chances of identifying a low number of equivalence classes, we considered a clustering technique based on modularity (*i.e.,* the Louvain method [241]), which allows reducing the amount of connections among groups of nodes belonging to different clusters, and, consequently, lowers the number of reference nodes to be considered for discovering equivalence classes.

Equivalence class with clustering

Let us assume a given graph $\mathbf{G}$ is split into a set of clusters $\mathbf{C}$, where a single cluster $\mathbf{C}_i$ is a connected sub-graph of $\mathbf{G}$ induced by a set of nodes $\mathbf{V}_{\mathbf{C}_i} \subset \mathbf{V}$.

For each cluster $\mathbf{C}_i \in \mathbf{C}$, it is possible to identify a set of *border nodes* $\mathbf{BN}_{\mathbf{C}_i}$. A border node $b_i \in \mathbf{BN}_{\mathbf{C}_i}$ is a node belonging to $\mathbf{C}_i$ and having at least one neighbor belonging to another cluster, as graphically presented in Fig. 5.7 (circled nodes are border nodes).

To discover equivalence classes, for each cluster $\mathbf{C}_i$, it is possible to group nodes based on their distance and number of shortest paths to the border nodes. To this end, we can leverage the following theorem (see [234] for the proof).

Theorem 5.4.1 *Let $k \in \mathbb{R}^+$ and $l \in \mathbb{R}$, let $\mathbf{C}_i$ be a generic cluster of graph $\mathbf{G}$ with border nodes $\mathbf{BN}_{\mathbf{C}_i}$, and $s, p \in \mathbf{V}_{\mathbf{C}_i}$. If $\forall\ b_j \in \mathbf{BN}_{\mathbf{C}_i}\ \sigma_{s,b_j} = k \cdot \sigma_{p,b_j}$ and $d_{\mathbf{G}}(s,b_j) = d_{\mathbf{G}}(p,b_j) + l$, then $\delta_{s,\overline{\mathbf{V}_{\mathbf{C}_i}}}(v) = \delta_{p,\overline{\mathbf{V}_{\mathbf{C}_i}}}(v)$, $\forall v \in \overline{\mathbf{V}_{\mathbf{C}_i}}$.*

In other words, any given pair of nodes s, p belonging to the sub-graph induced by nodes in cluster $\mathbf{C}_i$ (*i.e.,* $s, p \in \mathbf{V}_{\mathbf{C}_i}$), produces the same dependency score on all nodes $v \in \overline{\mathbf{V}_{\mathbf{C}_i}}$ for destinations $t \in \overline{\mathbf{V}_{\mathbf{C}_i}}$ if the distances and the number of shortest paths from s and p to every border node of $\mathbf{C}_i$ are the same, except for an additive or multiplicative factor, respectively.

From the previous theorem, we can derive the following corollary:

Corollary 5.4.1 *If $\forall\ b_j \in \mathbf{BN}_{\mathbf{C}_i}$: $\hat{\sigma}_{s,b_j} = \hat{\sigma}_{p,b_j}$ and $\hat{d_{\mathbf{G}}}(s,b_j) = \hat{d_{\mathbf{G}}}(p,b_j)$, then $\delta_{s,\overline{\mathbf{V}_{\mathbf{C}_i}}}(v) = \delta_{p,\overline{\mathbf{V}_{\mathbf{C}_i}}}(v)$, $\forall v \in \overline{\mathbf{V}_{\mathbf{C}_i}}$.*

where $\hat{d}_{\mathbf{G}}(s,b_j)$ represents the normalised distance of the generic node s to the generic border node b_j, defined as follows:

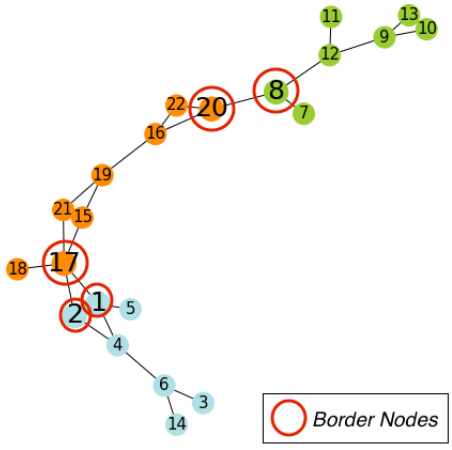
$$\hat{d}_{\mathbf{G}}(s,b_j) = d_{\mathbf{G}}(s,b_j) - min_{b_k \in \mathbf{BN}_{\mathbf{C}_i}} d_{\mathbf{G}}(s,b_k)$$

and $\hat{\sigma}_{s,b_j}$ represents the normalised number of shortest paths from the generic node s to the generic border node b_j, and is defined as:

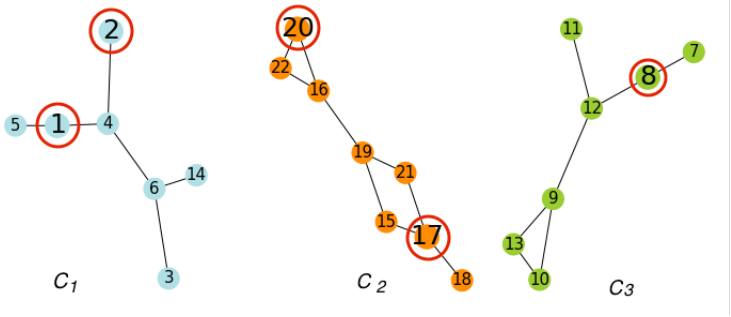
$$\hat{\sigma}_{s,b_j} = \sigma_{s,b_j} / min_{b_k \in \mathbf{BN}_{\mathbf{C}_i}} \sigma_{s,b_k}$$

Normalised distance and normalised shortest paths simplify the identification of classes since they are characterised by the same vector of normalised distances and shortest paths as explained below with an example and the support of a graphical representation.

Let $\mathbf{G}$ be the simple graph reported in Fig. 5.7(a), decomposed in three clusters, each separately shown in Fig. 5.7(b). We focus on the blue cluster, denoted as $\mathbf{C}_1$, in order to illustrate the concept of equivalence class (see Tab. 5.6 and Fig. 5.8).



(a) Example of clustered graph



(b) Three sub-graphs originated by clustering the graph

Figure 5.7: Example of clustering.

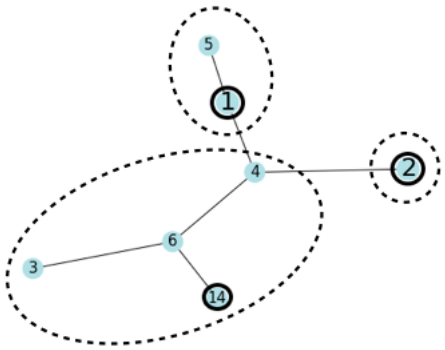


Figure 5.8: Classes of equivalent nodes in the blue cluster C_1

In $\mathbf{C}_1$, nodes **1** and **2** are border nodes (also denoted as b_1 and b_2 in Tab. 5.6). The nodes of $\mathbf{C}_1$ are related to b_1 and b_2 according to the properties detailed in Tab. 5.6: for each node the normalised distances and normalised number of shortest paths to the border nodes are reported. According to our previous definitions, nodes **3**, **4**, **6**, **14** and **5**, **1** can be grouped in two classes respectively, whereas node **2** is assigned to a singleton class. Nodes **1**, **2** and **14** are the pivots[16].

[16]In our previous version of the algorithm, the pivots were chosen to minimise the error. Here, as we will explain later, any node in a class can be a pivot.

node v	$\hat{d}_{C_1}(v, b_1)$	$\hat{d}_{C_1}(v, b_2)$	$\hat{\sigma}_{vb_1}$	$\hat{\sigma}_{vb_2}$
1	0	2	1	2
2	2	0	2	1
3	0	0	1	1
4	0	0	1	1
5	0	2	1	2
6	0	0	1	1
14	0	0	1	1

Table 5.6: Normalised distances and normalised number of shortest paths for the blue cluster C_1. Note that the values of $\hat{d}$ and $\hat{\sigma}$ are computed considering external node 17, as from Fig. 5.9(b).

Cluster-based exact BC computation

The equivalence classes allow us to compute the dependency score on nodes - and for destinations - that do not belong to the same cluster of the source node, which means that the contributions computed via this approach are only *partial*. To obtain the total BC, we rewrite Eq. 5.4 as follows:

$$
\begin{aligned}
BC(v) &= \sum_{s \in \mathbf{V}} \delta_{s,\bullet}(v) \\
&= \sum_{s \in \mathbf{V}} \sum_{t \in \mathbf{V}_{\mathbf{C}(s)}} \delta_{s,t}(v) + \sum_{s \in \mathbf{V}} \sum_{t \notin \mathbf{V}_{\mathbf{C}(s)}} \delta_{s,t}(v) \\
&= \underbrace{\sum_{s \in \mathbf{V}_{\mathbf{C}(v)}} \sum_{t \in \mathbf{V}_{\mathbf{C}(v)}} \delta_{s,t}(v)}_{\text{sum of local dependency scores} = \delta^{\lambda}(v)} + \underbrace{\sum_{s \in \mathbf{V}} \sum_{t \notin \mathbf{V}_{\mathbf{C}(s)}} \delta_{s,t}(v)}_{\text{sum of global dependency scores} = \delta^{\gamma}(v)} \\
&+ \underbrace{\sum_{s \notin \mathbf{V}_{\mathbf{C}(v)}} \sum_{t \in \mathbf{V}_{\mathbf{C}(s)}} \delta_{s,t}(v)}_{\text{sum of dependency scores on external nodes} = \delta^{\epsilon}(v)}
\end{aligned}
\tag{5.7}
$$

In Eq. 5.7, the term $\sum_{s \in \mathbf{V}} \sum_{t \in \mathbf{V}_{\mathbf{C}(s)}} \delta_{s,t}(v)$ is decomposed in two terms, by splitting the external sum operation (*i.e.,* $\sum_{s \in \mathbf{V}}$) into two parts, (*i.e.,* $s \in \mathbf{V}_{\mathbf{C}(v)}$ and $s \notin \mathbf{V}_{\mathbf{C}(v)}$), given that $t \in \mathbf{V}_{\mathbf{C}(v)}$ is equivalent to $t \in \mathbf{V}_{\mathbf{C}(s)}$, when $s \in \mathbf{V}_{\mathbf{C}(v)}$.

As a result, we can distinguish two main components in Eq. 5.7, *local* and *global* dependency scores. The third additional term is necessary to properly take into account the possible existence of shortest paths connecting nodes of the same cluster via nodes belonging to one or more different clusters, *i.e., external nodes*.

We define the *local* dependency score of a node s on a node v, $\delta^{\lambda}_{s,\bullet}(v)$, as the sum of pair dependency scores for which source s, the destinations and node v belong all to the same cluster. We define the *local BC* of a node v, $\delta^{\lambda}(v)$, as the BC of v computed on the sub-graph $\mathbf{C}(v)$.

Local BC is computed using Brandes' algorithm inside each cluster[17], which generates, as a by-product, additional information (*i.e.,* the number of shortest paths and distances to border nodes). This information is later used to group nodes into equivalence classes and to fasten the computation of global dependency scores.

The *global* dependency score of a node s on a node v, $\delta^{\gamma}_{s,\bullet}(v)$, is the sum of all the pair dependency scores for which destinations do not belong to the same cluster of source node s. The *global BC* of the generic node v, $\delta^{\gamma}(v)$, is thus the sum of the global dependency scores for source node s ranging over the whole set of nodes $\mathbf{V}$.

The dependency score of a node s on an external node v, *i.e.* $\mathbf{C}(v) \neq \mathbf{C}(s)$, noted as $\delta^{\epsilon}_{s,\bullet}(v)$, is the sum of all the pair dependency scores for which destinations belong to the

[17] As explained later, in the special case where there are external shortest paths in the cluster, the local BC is actually computed inside the *extended* cluster.

same cluster of the source node s. We denote by $\delta^{\epsilon}(v)$ the sum of all the dependency scores on v, when v is an external node and the sources and destinations are in the same cluster, different from $\mathbf{C}(v)$.

This last term $\delta^{\epsilon}(v)$ is equal to zero when the clustering is ideal, *i.e.* when all the shortest paths between any pair of nodes of a cluster only contain nodes from that same cluster. When this condition is not fulfilled, multiple side effects due to the presence of external nodes have to be taken into account, as discussed below.

External nodes/shortest paths

Given a cluster $\mathbf{C}_i$, two nodes $s, t \in \mathbf{C}_i$ and two border nodes $b_1, b_2 \in \mathbf{C}_i$, there may exist shortest paths between s and t that exit $\mathbf{C}_i$ through b_1, cross a certain number of nodes belonging to other clusters and then re-enter $\mathbf{C}_i$ through b_2. We call these shortest paths *external* shortest paths and the nodes lying on them that do not belong to $\mathbf{C}_i$, $\mathbf{EN}_{\mathbf{C}_i}$, *external* nodes of $\mathbf{C}_i$. If the existence of such external shortest paths is neglected, BC computation will be affected by an error due to the incorrect values of the lengths and the counts of shortest paths between pairs of nodes inside the same cluster. Consequently, an error in the computation of the local BC, δ^{λ}, and in the identification of equivalence classes will be introduced. This was one of the approximation errors that affected the previous versions of our algorithm [237, 232]. To remove this intra-cluster error, we borrowed the idea proposed by the authors in [239]. After clustering, we build a Hierarchical Sub-Network (HSN), *i.e.,* a sub-graph of $\mathbf{G}$ induced by the border nodes of all the clusters and nodes lying on the intra-cluster shortest paths between pairs of border nodes of the same cluster.

By retrieving all the shortest paths between pairs of border nodes of the same cluster via the HSN, we can identify possible external nodes for that cluster. Afterwards, we can extend each cluster with the related external nodes and use the extended clusters as sub-graphs to identify equivalence classes and pivots. Thus, local BC δ^{λ} can be correctly computed inside these extended clusters instead of the initial ones. The computation of local BC on the extended clusters also allows, as a by-product, computing the pair dependency scores on the external nodes. By accumulating the contributions on each external node from different clusters (in case a node is external to multiple clusters), it becomes trivial to compute the $\delta^{\epsilon}(v)$ term in Eq. 5.7.

Formally, an extended cluster $\mathbf{C}_i^*$ of a cluster $\mathbf{C}_i \in \mathbf{C}$ is defined as a connected sub-graph induced by nodes $\mathbf{V}_{\mathbf{C}_i^*} = \mathbf{V}_{\mathbf{C}_i} \cup \mathbf{EN}_{\mathbf{C}_i}$.

(a) Hierarchical Sub-Network from clusters in Fig. 5.7(b)

(b) Cluster $\mathbf{C}_1$, extended with the external node **17**

Figure 5.9: Example of external node found through the HSN.

We provide an illustrative example to better understand how the HSN is built and how it is used to form the extended clusters. Let us consider again the clustered graph from Fig. 5.7. In cluster $\mathbf{C}_1$, nodes **1** and **2** are border nodes, while node **4** lies on the only intra-cluster shortest path between them. In cluster $\mathbf{C}_2$, nodes **17** and **20** are border nodes and nodes **15**, **21**, **19** and **16** lie on the intra-cluster shortest paths between them. Finally,

in cluster $\mathbf{C}_3$, there is only border node **8**. All the aforementioned nodes build up the HSN (see Fig. 5.9(a)). If we now consider the shortest paths between border nodes **1** and **2** via the HSN, we notice that node **17** lies on a shortest path connecting the two former nodes. Consequently, it represents an external node of $\mathbf{C}_1$ (see Fig. 5.9(b)).

Dependency score of pivots

From the equivalence class relationship described in Sec. 5.4.3, a pivot of such a class only represents the dependency scores on nodes v - and for destinations t - which do not belong to its own cluster. In fact, given a cluster $\mathbf{C}_i \in \mathbf{C}$ and all its equivalence classes $\mathbf{K}_{\mathbf{C}_i}$, from Eq. 5.6, we have:

$$\sum_{s \in \mathbf{K_i}} \delta_{s,\overline{\mathbf{V}_{\mathbf{C}_i}}}(v) = |\mathbf{K}_i| \cdot \delta_{k_i,\overline{\mathbf{V}_{\mathbf{C}_i}}}(v) \quad \forall v \in \overline{V_{\mathbf{C}_i}}, \mathbf{K}_i \in \mathbf{K}_{\mathbf{C}_i}. \tag{5.8}$$

This equation can be exploited to speed up the computation of BC, by building on Brandes' algorithm and SSSP explorations, but only holds if $v \in \overline{\mathbf{V}_{C_i}}$. Thus, it cannot be directly applied to correctly compute values of global BC when v is in the same cluster of the source. Therefore, the algorithm requires a more elaborated approach to properly and efficiently calculate the contribution from the pivot of $\mathbf{K}_{\mathbf{C}_i}$ to the BC of nodes $v \in \mathbf{V}_{C_i}$[18].

First of all, let us decompose the global dependency scores from Eq. 5.7 based on the cluster of node v as follows:

$$\delta^{\gamma}(v) = \sum_{s \notin \mathbf{V}_{\mathbf{C}(v)}} \sum_{t \notin \left(\mathbf{V}_{\mathbf{C}(v)} \cup \mathbf{V}_{\mathbf{C}(s)}\right)} \delta_{s,t}(v) + \sum_{s \notin \mathbf{V}_{\mathbf{C}(v)}} \sum_{t \in \mathbf{V}_{\mathbf{C}(v)}} \delta_{s,t}(v) + \sum_{s \in \mathbf{V}_{\mathbf{C}(v)}} \sum_{t \notin \mathbf{V}_{\mathbf{C}(v)}} \delta_{s,t}(v) \tag{5.9}$$

The previous equation can be further simplified by considering the following claim (see [234] for the proof):

Claim 5.4.1 *In undirected graphs:*

$$\sum_{s \in \mathbf{V}_{\mathbf{C}(v)}} \sum_{t \notin \mathbf{V}_{\mathbf{C}(v)}} \delta_{s,t}(v) = \sum_{s \notin \mathbf{V}_{\mathbf{C}(v)}} \sum_{t \in \mathbf{V}_{\mathbf{C}(v)}} \delta_{s,t}(v) \tag{5.10}$$

By relying on Eq. 5.10, it becomes possible, for any given node v, to replace with zero the sum of the pair-dependencies $\delta_{s,t}(v)$ for which $s \in \mathbf{V}_{\mathbf{C}(v)}$ and $t \in \overline{\mathbf{V}_{\mathbf{C}(v)}}$ (the third term in Eq. 5.9) and compensate later the lack of this term by doubling the sum of the pair-dependencies $\delta_{s,t}(v)$ for which $s \in \overline{\mathbf{V}_{\mathbf{C}(v)}}$ and $t \in \mathbf{V}_{\mathbf{C}(v)}$ (the second term in Eq. 5.9). In other words, Eq. 5.9 can be rewritten for undirected graphs as follows:

$$\delta^{\gamma}(v) = \sum_{s \notin \mathbf{V}_{\mathbf{C}(v)}} \sum_{t \notin \left(\mathbf{V}_{\mathbf{C}(v)} \cup \mathbf{V}_{\mathbf{C}(s)}\right)} \delta_{s,t}(v) + 2 \cdot \sum_{s \notin \mathbf{V}_{\mathbf{C}(v)}} \sum_{t \in \mathbf{V}_{\mathbf{C}(v)}} \delta_{s,t}(v) \tag{5.11}$$

With this formulation, the computation of global BC for any given node v can be performed by only considering the contributions that source nodes have on nodes v and for destinations t, where both v and t are outside the source's cluster. This satisfies the requirement for applying equivalence classes to compute dependency scores. Therefore, we can now use pivots as sources to efficiently compute the exact global BC.

In particular, let $\delta^{\gamma}_{s,\mathbf{V}_{\mathbf{C}(v)}}(v)$ and $\delta^{\gamma}_{s,\overline{\mathbf{V}_{\mathbf{C}(v)}}}(v)$ be the global dependency scores from node s on node v for destinations not belonging to $\mathbf{C}(s)$, but belonging to $\mathbf{C}(v)$, and the global

[18] In our previous version of the algorithm, we used Eq. 5.8 without taking into account the cluster of v and the pivots were chosen to minimize the error. Here, we avoid such an error during the computation of global dependency scores by exploiting the properties of undirected graphs, as explained later.

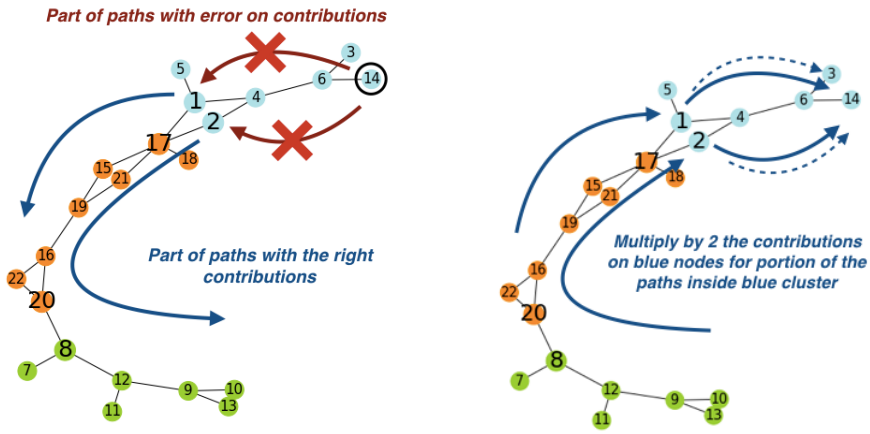
dependency score from node s on node v for destinations not belonging to $\mathbf{C}(s)$ and $\mathbf{C}(v)$. Eq. 5.11 can be rewritten as follows:

$$\delta^{\gamma}(v) = \sum_{s \notin \mathbf{V}_{\mathbf{C}(v)}} [2 \cdot \delta^{\gamma}_{s,\mathbf{V}_{\mathbf{C}(v)}}(v) + \delta^{\gamma}_{s,\overline{\mathbf{V}_{\mathbf{C}(v)}}}(v))] \tag{5.12}$$

Therefore, given a cluster $\mathbf{C}_i \in \mathbf{C}$ and all its equivalence classes $\mathbf{K}_{\mathbf{C}_i}$, we have:

$$\begin{aligned} &\forall v \notin \mathbf{V}_{\mathbf{C}_i}, \mathbf{K}_i \in \mathbf{K}_{\mathbf{C}_i}, \\ &\sum_{s \in \mathbf{K}_i} \left(2 \cdot \delta^{\gamma}_{s,\mathbf{V}_{\mathbf{C}(v)}}(v) + \delta^{\gamma}_{s,\overline{\mathbf{V}_{\mathbf{C}(v)}}}(v)\right) = |\mathbf{K}_i| \cdot \left(2 \cdot \delta^{\gamma}_{k_i,\mathbf{V}_{\mathbf{C}(v)}}(v) + \delta^{\gamma}_{k_i,\overline{\mathbf{V}_{\mathbf{C}(v)}}}(v)\right) \end{aligned} \tag{5.13}$$

Eq. 5.13 means that, during the back-propagation phase, we should distinguish between contributions due to destinations inside the same cluster of v and contributions due to destinations outside the cluster of v (and the cluster of s).



(a) Problem with nodes belonging to clusters of pivots (*e.g.,* 1, 2). Dependency scores from node 14 are set to null.

(b) Solution for undirected graphs. The contributions on nodes of the blue cluster (*e.g.,* 1, 2) are corrected by doubling dependency scores from sources of other clusters for destinations inside the blue cluster).

Figure 5.10: Global SSSP explorations from pivots

For a better understanding of the formulas above, let us consider an illustrative example by leveraging again the clustered graph from Fig. 5.7 and the equivalence classes of cluster $\mathbf{C}_1$ from Fig. 5.8.

The pivot node of the equivalence class composed of nodes {**3, 4, 6, 14**} is node **14**. According to the proposed approach, we calculate the dependency scores from node **14** on all nodes of clusters $\mathbf{C}_2$ (orange) and $\mathbf{C}_3$ (green) and multiply them by 4, avoiding to calculate the dependencies scores from nodes **3, 4** and **6**. This way, the computation time is divided by 4. However, while it is correct to multiply by 4 the dependency scores for nodes in $\mathbf{C}_2$ and $\mathbf{C}_3$, it is not for nodes belonging to the same cluster of the pivot (see Fig. 5.10(a)) since nodes **14, 3, 4, 6** of the class are indeed equivalent, but only with respect to border nodes of cluster $\mathbf{C}_1$ (*i.e.,* nodes **1**, **2**). Therefore, we cannot multiply by 4 the dependency scores on the nodes internal to cluster $\mathbf{C}_1$ (*i.e.,* **1, 2, 3, 4, 5, 6**) since these scores are not the same when computed, for instance, from node **14** or node **4**. To avoid the problem, we put these dependency scores to 0 during the SSSP explorations performed from pivot node **14**. These neglected scores will be later compensated (via multiplication by two) during the SSSP explorations performed from the pivot nodes of $\mathbf{C}_2$ and $\mathbf{C}_3$ (see Fig. 5.10(b)).

Back-propagation. Differently from Brandes' algorithm, it is not possible to directly express the global dependency score[19] of a node v, $\delta^{\gamma}_{s,\bullet}(v)$, in terms of the global dependency scores of w, $\delta^{\gamma}_{s,\bullet}(w)$, where $v \in \mathbf{P}_s(w)$. Indeed, when $\mathbf{C}(v) \neq \mathbf{C}(w)$ (*i.e.,* when crossing a cluster), the set of destinations of w which do not belong to $\mathbf{C}(w)$ can be composed of both destinations belonging to $\mathbf{C}(v)$ and destinations not belonging to $\mathbf{C}(v)$: for the former, the pair-dependencies have to be multiplied by 2, whereas for the latter no further operation is needed (see Eq. 5.13).

To overcome this problem, we decided to apply the classic recursive formula of Brandes' algorithm (Eq. 5.5) on a *vector* of contributions while propagating the global dependency scores $\delta^{\gamma}_{s,\bullet}(v)$. The dimensions of this vector of contributions correspond to the number of clusters, so that the contribution due to a destination t is assigned to $\delta^{\gamma}_{s,\mathbf{V}_{\mathbf{C}(t)}}(v)$. Formally, we have the following recursive formula:

$$\forall\, \mathbf{C}_i \in \mathbf{C} \setminus \mathbf{C}(s) : \delta^{\gamma}_{s,\mathbf{V}_{\mathbf{C}_i}}(v) = \sum_{w:v\in\mathbf{P}_s(w)} \frac{\sigma_{s,v}}{\sigma_{s,w}} * (\mathbb{1}_{w\in\mathbf{C}_i} + \delta^{\gamma}_{s,\mathbf{V}_{\mathbf{C}_i}}(w)), \tag{5.14}$$

where $\mathbb{1}_{w\in\mathbf{C}_i}$ represents a Boolean variable equal to 1 if $w \in \mathbf{C}_i$, 0 otherwise[20]. At the end of the back-propagation phase, we put the dependency scores of nodes v belonging to the same cluster of the (pivot) source node to 0, whereas the dependency scores of nodes belonging to the other clusters are computed using the following formula:

$$\delta^{\gamma}_{s,\overline{\mathbf{V}_{\mathbf{C}(s)}}}(v) = 2 \cdot \delta^{\gamma}_{s,\mathbf{V}_{\mathbf{C}(v)}}(v) + \sum_{\mathbf{C}_i\neq\mathbf{C}(v)} \delta^{\gamma}_{s,\mathbf{V}_{\mathbf{C}_i}}(v) \tag{5.15}$$

Finally, according to Eq. 5.13, $\delta^{\gamma}_{s,\bullet}(v)$ is multiplied by the cardinality of the equivalence class s belongs to.

5.4.5 Empirical Evaluation with Synthetic and Realistic Networks

We compared the execution times obtained with the implementation of our algorithm (named *E1C-FastBC*) to those obtained with other algorithms by using the Algorithmic Speedup (AS). Given two algorithms, a_1 and a_2, the AS of a_1 over a_2 with p cores, noted as $AS_p^{a_1/a_2}$, is defined as $T_p^{a_2}/T_p^{a_1}$, where $T_p^{a_2}$ and $T_p^{a_1}$ are the computation times obtained with p cores using algorithms a_2 and a_1, respectively. Hence, the larger the value of AS, the faster a_1 is compared to a_2 with the same computing resources. For example, $AS_p^{a_1/a_2} = 2$ means that the time taken by a_1 is half the time taken by a_2 and therefore that a_1 is two times faster than a_2.

In particular, we compared the *E1C-FastBC* algorithm, labelled with $\mathcal{E}$, with Brandes' algorithm, labelled as $\mathcal{B}$ and with the solution proposed in [239], labelled with $\mathcal{H}$. We chose this algorithm for comparison because it belongs to the same category as ours (cluster-based computation) and it addresses the problem of exact BC computation.

However, due to the unavailability of source/executable code for $\mathcal{H}$, we only considered the AS metric in sequential mode, by relying on the indications provided by the authors in the paper for its computation (see Eq. 7 in [239]).

To further explore the performance of our solution, we also analysed the efficiency of the *E1C-FastBC* algorithm, based on the canonical definition of speedup. Specifically, the speedup obtained with p cores is defined as $S_p = T_s/T_p$, where T_s is the computation time in sequential mode and T_p is the computation time with p cores. The efficiency with p cores, noted as E_p, is then defined as S_p/p[21].

[19] $\delta^{\gamma}_{s,\bullet}(v)$ is equivalent to $\delta_{s,\overline{\mathbf{V}_{\mathbf{C}(s)}}}(v)$

[20] This is the part of contribution due to w as a destination

[21] Efficiency E_p is strictly, but non linearly, related to the AS metric when used to compare to different algorithms. From the definition of speedup: $T_p^{a_2} = T_s^{a_2}/S_p^{a_2}$, where $T_s^{a_2}$ and $S_p^{a_2}$ are the execution time

In all reported tests, we checked the accuracy of our solution by always observing zero error on BC values.

Datasets and Experimentation Testbed

In our experimental tests, we considered both synthetic and real graphs.

For the first category, we focused on scale-free graphs generated using the implementation of the Barabási-Albert model provided by the Python library NetworkX. According to that model, a graph of n nodes is grown by attaching new nodes, one at a time, each with m' edges that are preferentially attached to existing nodes with high degree. In our case, m', which is called the preferential attachment coefficient, was chosen equal to 1. This led us to consider graphs with $m = n - 1$ edges and an average degree approximately equal to 2[22]. This choice was motivated by the features of the current implementation of our algorithm that benefits of high modularity. In other words, this class of dataset was considered as the best-case scenario. However, as mentioned in the introduction, this does not limit the applicability of our solution because many real-world systems can be represented with the Barabási-Albert model. In particular, to analyse the algorithm in terms of performance and scalability, we generated graphs with different sizes (see Table 5.7).

	Graph	n	m	d_{avg}	d_{max}	cc_{avg}
Synthetic	barabási-albert	6,250	6,249	1.999	126	0.000
	barabási-albert	12,500	12,499	"	225	"
	barabási-albert	25,000	24,999	"	344	"
	barabási-albert	50,000	49,999	"	463	"
	barabási-albert	100,000	99,999	"	1,138	"
	barabási-albert	200,000	199,999	"	676	"
	barabási-albert	400,000	399,999	"	1,142	"
	barabási-albert	800,000	799,999	"	1,587	"
Real	web-webbase-2001 [242]	16,062	25,593	3.187	1,679	0.224
	ego-twitter [243]	22,322	31,823	2.851	238	0.072
	internet [242]	124,651	193,620	3.107	151	0.062
	lyon-road-network	156,102	178,845	2.291	8	0.017
	email-euAll [244]	224,832	339,925	3.024	7,636	0.079

Table 5.7: Topological information of synthetic & real graphs. The names of the graphs are given in the first column, whereas the number of nodes and edges are given in the second and third columns. d_{avg} and d_{max} are the average and max degree, respectively. cc_{avg} is the average clustering coefficient.

For the second category, we focused on real graphs[23] available in public datasets. Table 5.7 reports all the graphs used, together with some relevant properties. In particular, for each graph, the table reports the average degree (d_{avg}), the max degree (d_{max}) and the average clustering coefficient (cc_{avg}).

All the datasets, except the one related to the Lyon road network[24], are scale-free graphs. For the sake of brevity, we only report results related to the synthetic graphs and the lyon-road-network dataset in this document.

in sequential mode and the speedup obtained with p cores of algorithm a_2. Similarly, $T_p^{a_1} = T_s^{a_1}/S_p^{a_1}$. By these equations, the AS metric can be rewritten as: $AS_p^{a_1/a_2} = (T_s^{a_2}/T_s^{a_1}) \cdot (S_p^{a_1}/S_p^{a_2})$. Since for a given number of cores p, the ratio of the speedups equals the ratios of efficiency of the two algorithms, *i.e.*, $S_p^{a_1}/S_p^{a_2} = E_p^{a_1}/E_p^{a_2}$, we thus have $AS_p^{a_1/a_2} = (T_s^{a_2}/T_s^{a_1}) \cdot (E_p^{a_1}/E_p^{a_2})$. The relationship between AS and E_p suggests that if a_1 and a_2 have comparable efficiency with p cores, then the AS only depends on the ratio of the execution times of the two algorithms in sequential mode, thus providing interesting insights to analyse the two different solutions comparatively.

[22] Each link contributes 2 to the sum of the degrees. Thus, the sum of all degrees in the network is $2 \cdot (N - 1)$. For large values of N, the average degree approaches 2.

[23] For each graph, we extracted the largest connected component. Then, the latter was converted in an unweighted undirected graph.

[24] The lyon-road-network dataset was supplied by the French National Institute of Geographic Information (IGN). https://www.ign.fr

The tests were conducted on the PROMENADE-v2.0 platform prototype (Sec. 5.3.1), using the Spark BP component for our experiments. It was deployed on a server machine with 128 GB of RAM and 2 sockets Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz, with 14 physical cores and 2 threads per core for a total of 28 logical cores per socket and 56 virtual cores in hyper threading, running Linux Debian as operating system. Both Brandes' algorithm and *E1C-FastBC* were implemented in Scala.

Synthetic Graphs Analysis

Fig. 5.11 shows the algorithmic speedup of *E1C-FastBC* over Brandes' algorithm, $AS_p^{\mathcal{E}/\mathcal{B}}$, obtained on the synthetic graphs in both sequential and parallel modes. In particular, we doubled the number of nodes from 25,000 to 800,000, and we considered a number of cores p equal to 1, 5, 10, 15, 20 and 25. We estimated by log-log regression the computation times with Brandes' algorithm for the graphs with 400,000 and 800,000 nodes, since executions would have required weeks to complete, whereas our algorithm completed in maximum 31.5 minutes and 1.64 hours, respectively (sequential mode).

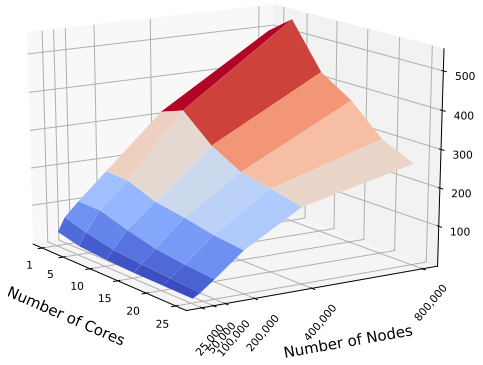
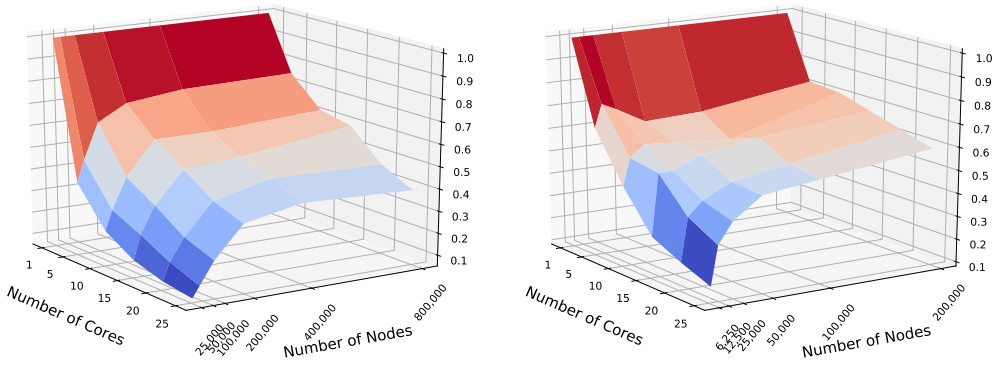


Figure 5.11: Comparison with Brandes' algorithm. Algorithmic speedup analysis - $AS_{p=[1,5,10,15,20,25]}^{\mathcal{E}/\mathcal{B}}$

As shown in Fig. 5.11, $AS_p^{\mathcal{E}/\mathcal{B}}$ increases with the size of the graph, meaning that *E1C-FastBC* is not only faster than Brandes' algorithm but its speedup grows with larger graphs. This is due to the fact that the computation of our algorithm is strongly dependent on the number of border nodes ($|\mathbf{BN}|$), pivots ($|\mathbf{K}|$) and external nodes ($|\mathbf{EN}|$), in addition to the number of nodes (n) and edges (m). The first two variables increase slowly compared to the number of nodes and edges, while the third is almost always zero (only in one case it was equal to 2). The drawback is that $AS_p^{\mathcal{E}/\mathcal{B}}$ decreases as the number of cores increases. This behaviour is due to the fact that the Brandes' algorithm is more efficient than *E1C-FastBC* (see Fig. 5.12).

This means that the ratio $E_p^{a_1}/E_p^{a_2}$ in the relationship between the AS metric and the efficiency is lower than 1. Consequently, the AS value in the sequential case is not preserved as the number of cores increases. However, as the following efficiency analysis will further clarify, this does not mean that *E1C-FastBC* is less scalable than Brandes' algorithm, but rather that it needs very large graphs to better exploit the available computing resources. This statement is also confirmed by Fig. 5.11, which clearly shows that when the graph size is 400,000, a higher number of cores performs even better than a smaller one: in particular, we have that the $AS_p^{\mathcal{E}/\mathcal{B}}$ is better with 5 cores than with 1 core. To have a similar behaviour even for a number of cores greater than 5, we should consider larger graphs.

To better understand the performance of *E1C-FastBC*, we investigated its efficiency with respect to that of Brandes' algorithm. Fig. 5.12 reports the results of the efficiency analysis



(a) E1C-FastBC algorithm

(b) Brandes' algorithm

Figure 5.12: Comparison with Brandes' algorithm. Efficiency analysis - $E_{p=[1,5,10,15,20,25]}$

performed for the two algorithms. In both cases, it is possible to observe that: *(i)* the efficiency decreases as the number of cores increases and *(ii)* for a given number of cores, it increases as the number of nodes increases.

However, it is worth highlighting that in the efficiency analysis, we use different but overlapping ranges of values for the number of nodes. In particular, for our solution, we select larger graphs since we aim to show that our algorithm scales well, especially with very large graphs. In fact, the efficiency trend is almost the same in the two cases reported in Fig. 5.12. Moreover, given the maximum values of the number of nodes for the two algorithms (800,000 for ours, 200,000 for Brandes'), efficiency values are approximately the same with 5 cores (*i.e.,* the first considered parallel configuration) but significantly diverge as the number of cores increases. In particular, efficiency of *E1C-FastBC* decreases with a higher rate than that of Brandes.

The reason for this behaviour lies in the reduced amount of computation required by our solution, which can be satisfied with limited parallelism (*i.e.,* number of cores). Indeed, pivots allow to significantly decrease the number of (modified) Brandes' SSSP explorations performed on the whole graph, which represent the heaviest part of the whole computation, thus reducing the workload of each core.

Our solution also introduces another benefit: it allows to mitigate the variability of the computation times due to the different topological characteristics of the graphs and to the partitioning of data performed by Spark during executions. Indeed, there may exist some partitions of the RDDs characterized by a high concentration of nodes that generate the most complex shortest path trees. The time required to process these partitions directly impacts the time required to process the whole RDD, since partitions are processed in parallel. However, Spark tasks process each single partition sequentially. This aspect, combined with the fact that the number of partitions of an RDD is always equal to the number of cores and the default partitioning scheme of Spark distributes data evenly across the partitions, explains the punctual efficiency drops that can be observed in the plot related to Brandes' algorithm, when using graphs with 50,000 and 100,000 nodes and a low number of cores (*i.e.,* the cavity that can be observed for 50,000 and 100,000 nodes with 5-10 cores in the 3d plot of Fig. 5.12(b)).

Fig. 5.13(a) reports the algorithmic speedup of *E1C-FastBC* over Brandes' algorithm, $AS_{p=1}^{\mathcal{E}/\mathcal{B}}$, alongside with the algorithmic speedup of the approach in [239] over Brandes, $AS_{p=1}^{\mathcal{H}/\mathcal{B}}$, on synthetic graphs and in sequential settings. $AS_{p=1}^{\mathcal{H}/\mathcal{B}}$ is analytically computed based on Eq. 7 provided in [239].

Using such equation, it is possible to observe that: *(i)* $AS_{p=1}^{\mathcal{H}/\mathcal{B}}$ depends on the number

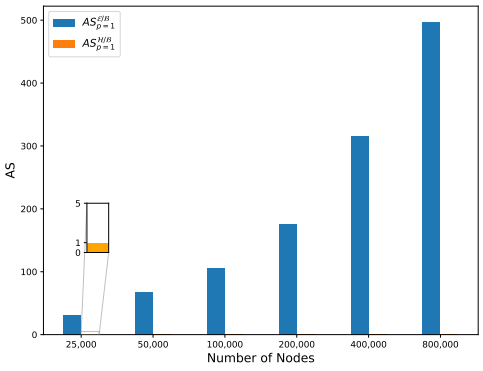
of clusters ($|\mathbf{C}|$) and the average degree (d_{avg}), and *(ii)* when $|\mathbf{C}| + 2 \gg d_{avg}/2$, it can be approximated with $d_{avg}/2$. Therefore, since for synthetic graphs the average degree is constant and the number of clusters increases with the number of nodes, $AS_{p=1}^{\mathcal{H}/\mathcal{B}}$ is always approximately equal to 1 (the average degree is 2). In particular, the higher the number of clusters, the closer to 1 the $AS_{p=1}^{\mathcal{H}/\mathcal{B}}$. This means that the algorithm proposed in [239] cannot improve that of Brandes. Conversely, ours improves Brandes by a large multiplicative factor. We can thus conclude that our solution always outperforms the one in [239] with synthetic graphs.

Real Graphs Analysis

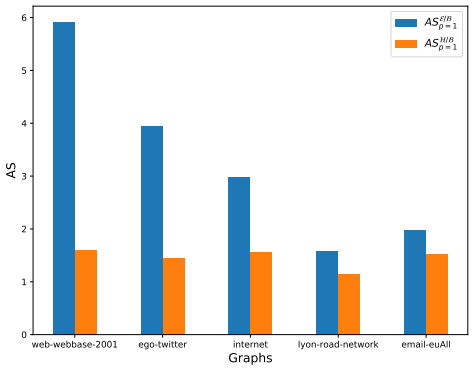
Fig. 5.13(b) reports the results of the analysis of $AS_{p=1}^{\mathcal{E}/\mathcal{B}}$ and $AS_{p=1}^{\mathcal{H}/\mathcal{B}}$ carried out on real graphs. $AS_{p=1}^{\mathcal{H}/\mathcal{B}}$ is computed again using Eq. 7 provided in [239]. In all cases, our solution outperforms the one in [239].

To further confirm the considerations on the scalability of our solution, reported in the previous section, we analyse in the following both the algorithmic speedup and efficiency values of *E1C-FastBC* on the *lyon-road-network* graph, for which we observed a very high number of pivots (about 60% of the number of nodes) and the lowest algorithmic speedup factor. As shown in Fig. 5.14(a), the AS is always greater than 1, thus confirming the usefulness of our solution, although the reported values are not comparable to those obtained on synthetic graphs (see Fig. 5.11) with a similar number of nodes (100,000 and 200,000). In spite of this, the algorithm becomes more scalable and efficient than in the case of synthetic graphs with 100,000 and 200,000 nodes due to the increased amount of computation resulting from the higher number of border nodes, pivots and external nodes (see Fig. 5.14(b)).

In conclusion, our performance analysis confirmed the possibility to use E1C-FastBC for faster computation of BC and motivated its integration in the PROMENADE-v2.0 platform core. Specifically, we observed computation times in the order of few minutes to few seconds for analysing a large road network (more than 150,000 nodes and 170,000 links) such as that of Lyon, France, depending on the available hardware, and with limited parallelism. These figures indicate that our solution can effectively support (quasi) real-time monitoring of large networks, thereby representing a useful tool in the context of resilience enhancement of rapidly varying networks.

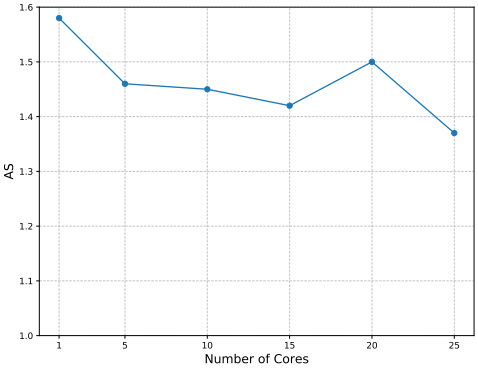


(a) Algorithmic speedup analysis with synthetic graphs in sequential mode - $AS_{p=1}^{\mathcal{E}/\mathcal{B}}$ and $AS_{p=1}^{\mathcal{H}/\mathcal{B}}$. $AS_{p=1}^{\mathcal{H}/\mathcal{B}}$ is always equal to 1 for synthetic graphs (see Subsection 5.4.5). We therefore add only one single zoomed bar in the figure just to make it visible to the reader.

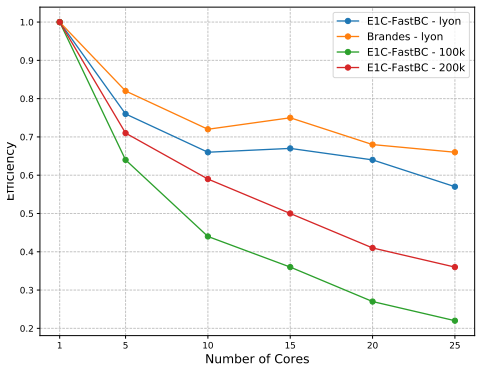


(b) Algorithmic speedup analysis with real graphs in sequential mode - $AS_{p=1}^{\mathcal{E}/\mathcal{B}}$ and $AS_{p=1}^{\mathcal{H}/\mathcal{B}}$

Figure 5.13: Comparison with the HSN solution for synthetic (a) and real graphs (b).



(a) Algorithmic speedup analysis for lyon-road-network graph in parallel mode - $AS^{\mathcal{E}/\mathcal{B}}_{p=[1,5,10,15,20,25]}$



(b) Efficiency analysis for lyon-road-network graph - $E_{p=[1,5,10,15,20,25]}$. Comparison with synthetic graphs having similar size (100k and 200k nodes) and with Brandes' algorithm on the same graph.

Figure 5.14: Performance analysis of the E1C-FastBC algorithm versus Brandes on the real graph lyon-road-network.

Moreover, our solution could benefit from customised clustering methods for more effective identification of border nodes in (synthetic and real) graphs with different topologies and further improve its scalability. The efficiency of E1C-FastBC could be also improved by means of a better mapping of the algorithm on distributed resources, when data-parallelism is exploited, and by improving locality especially when different Spark executors are used.

5.5 Multi-agent-based Large-Scale Traffic Control

As a second case study of the PROMENADE platform, we proposed a multi-agent macroscopic solution for large-scale traffic control in urban environments [245].

Traffic control strategies are fundamental to reducing congestion and pollution, but they require to address multiple challenges to be effective, especially if the goal is to address large traffic networks encompassing multiple and heterogeneous road axis and multi-modal mobility options. This is typical in urban areas of large metropolitan modern cities. Moreover, besides efficiency and scalability, the control strategy should respect user privacy by relying on a minimum to null amount of personal information, preserving accessibility by not penalising specific travellers or city areas, and introducing minimal burden to individuals with respect to the travel experience. Furthermore, large-scale control strategies should help achieve resilience objectives, *i.e.,* by reducing circulation in highly-vulnerable areas and enforcing effective rerouting in the presence of accidents or for emergency evacuation.

Traditional approaches to traffic control include perimeter control [246] and optimal route guidance [247]. Despite their proven effectiveness in controlled environments and some successful small-scale on-the-field application, such approaches usually suffer in real-world settings, either in terms of application to large metropolitan networks (*i.e.,* they tend to simply move congestion outside the controlled areas) or in terms of user acceptability as they depend completely upon the users following the rerouting suggestions. Moreover, such solutions typically rely on centralised approaches, which significantly reduces their scalability [248].

Based on these considerations, we designed and experimented, in the last part of Cecile Daniel's thesis, a novel decentralised control strategy, named COoperative Multi-agent FramewOrk for large-scale Routing-based Traffic control (COMFORT). The solution aims at homogeneously distributing road traffic between zones of a large urban network. The

idea is to periodically assess, in a real-time like fashion, the traffic conditions locally to each zone, by relying on traffic sensors (such as loop detectors) and the MFD theory [249]. Local information about traffic congestion is shared between zones via gossiping to guarantee rapid reaction to congestion formation and indirectly transmitted to vehicles through rerouting suggestions. The choice of being rerouted and the possible new route computation are performed at the vehicle level, ensuring privacy by design[25] and leaving the final choice of rerouting to the driver[26].

Differently from [248] and by fully adhering to the principles of the PROMENADE architecture in terms of scalability requirements, we adopted a fully decentralised approach. The latter is based on a cooperative Multi-Agent System (MAS) paradigm partially inspired by the work of Lequay *et al.* [250] related to reducing energy consumption in an electricity distribution network. We designed and implemented a simulation-based prototype of COMFORT as part of the PROMENADE platform.

According to the MAS paradigm, vehicles were considered in COMFORT as one specific kind of agent. Geographical zones were also considered as agents. They process information related to the accumulation of vehicle agents within the corresponding spatial region and accomplish specific interactions (mainly control actions, but also information sharing) with vehicles and neighbouring zone agents. This inherently defines a multi-layer framework. The main interactions between the two considered kinds of agents are reported in the activity diagram of Fig. 5.15.

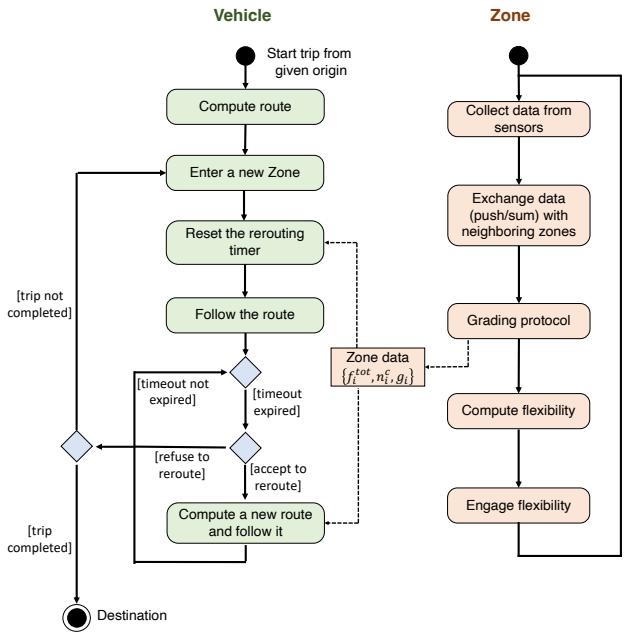


Figure 5.15: Activity diagram of the interactions between agents in COMFORT.

The zone agents are associated with attributes that represent the main time-dependent variables involved in the control strategy: *(i) total flexibility*, $f_i^{tot}(t)$, a time-dependent function inversely related to the congestion state of the zone, *i.e.,* the difference between the

[25] In our approach, vehicles do not communicate their final destination or the route they are following. Instead, they only signal the next area they are willing to traverse to the zone they are currently in.

[26] It is worth noting that COMFORT assumes a share of travelling vehicles to be connected via the urban infrastructure or via a software mobile application available to the driver. Connected vehicles are the only vehicles that can receive recommendations for rerouting from the area they are travelling through. On the contrary, traffic conditions (estimated via sensors) include both connected and non-connected vehicles. To model different scenarios of connectivity and user acceptability, we tested different values for the probability of a vehicle accepting a rerouting suggestion.

critical accumulation (derived from the MFD of the zone) and the current observation of traffic conditions in the zone[27]; *(ii) actual flexibility*, $f_i(t)$, representing the evolution of traffic accumulation in the zone, computed over consecutive discrete time steps and corresponding to the number of vehicles actually accepted by the given zone during the last time step; *(iii) grade*, $g_i(t)$, expressing the ability of the zone to cooperate with its neighbours by reliably engaging its flexibility over time, *i.e.,* by accepting a sufficiently high number of vehicles with respect to its total flexibility. The core concept of the control strategies is to generate rerouting advice for vehicles before they enter a congested area. This is achieved by computing a rerouting timeout, which depends on a combination of the following variables: the level of congestion (*i.e.,* lower total flexibility) and the reliability (*i.e.,* the grade) of the neighbouring zone. Essentially, the more congested and less reliable a given neighbouring zone is, the sooner a vehicle expected to enter it will be advised to compute an alternative route, avoiding that zone. A detailed description of the aforementioned variables and of the rerouting strategy can be found by the interested reader in [245].

The preliminary prototype of COMFORT has been evaluated through simulation on a grid Manhattan network with 3,712 nodes and 10,324 edges, partitioned into 25 (5X5) zones, enabling a comparison of our control strategy performance depending on the zone size and density. We considered the Total Travel Time Reduction ($\%TR$), expressed in percentage, to evaluate the impact of our strategy and compare the results to a no-control baseline:

$$\%TR = \frac{TTT_{nocontrol} - TTT_{control}}{TTT_{nocontrol}} \tag{5.16}$$

with TTT_i representing the total travel time, *i.e.,* the sum of the duration of all trips generated during the simulation, when using a specific control strategy i. A positive $\%TR$ indicates the control strategy has been capable of improving the reference no-control scenario, whereas a negative $\%TR$ points out a pejorative impact of the control strategy with respect to the reference situation. An additional element to be considered to evaluate the effectiveness of the control strategy is the number of generated reroutings. We hypothesised that rerouting represents a cost for the driver, and thus the strategy should be tuned to maximising $\%TR$ with a low number of vehicle rerouting.

In the Manhattan scenario, we measured a consistent improvement in terms of $\%TR$ over the reference no-control situation, with respect to different combinations of hyper-parameters. The best configuration allowed achieving a $\%TR$ up to 16%, for a total number of approximately 2 500 reroutings (with respect to 16, 743 vehicles considered in the simulation).

Similar analyses were performed also in the realistic case of the road network of Lyon, France (see Fig. 5.16). We observed an improvement in terms of $\%TR$ up to 18% in the best hyper-parameter configuration. Very few reroutings were necessary to improve the reference situation. The maximum percentage of rerouted vehicles was 6.7%, corresponding to 4, 700 rerouted vehicles.

The results reported in this section are encouraging and represent a first step towards an efficient and privacy-aware real-time traffic control solution for large-scale urban networks. Additional efforts are needed to refine the control strategy and prove the applicability and efficiency of the proposed solution. In particular, the prototype requires an extensive evaluation in terms of the actual acceptability rate of the rerouting suggestions by travellers as well as the critical amount of connected vehicles necessary to significantly reduce the overall urban congestion. In that respect, we are considering a real-world deployment of the COMFORT prototype on the PROMENADE platform, which would require the real-time collection of traffic flow data, jointly with the development of a mobile application to provide a real-time rerouting service to vehicles. We are also currently exploring the potential integration of the proposed approach with our existing solutions for efficiently computing

[27] A zone has a high flexibility when it can accept a high number of vehicles with respect to its critical capacity.

Figure 5.16: The northern region network of the conurbation of Lyon (3rd, 6th district of Lyon and Villeurbanne). The network contains 1,883 nodes, 3,383 links and is divided in 17 zones. Each zone is represented by a distinct colour.

betweenness centrality. The primary idea involves incorporating an additional term, representing a time-varying zone-criticality indicator, into the rerouting trigger. This indicator is computed by aggregating the BC values of nodes within neighbouring zones. The aim is to mitigate the pressure on areas that could become more prone to grid-locks at any given moment due to a high traffic concentration and limited alternative routes available to bypass emerging traffic bottlenecks or capacity drops (*e.g.,* due to accidents or other disruptive events). Lastly, to further minimise reaction time and introduce proactive measures in managing traffic congestion and disruptions, we are working on developing traffic forecasting solutions (see Chapter 6) that could be integrated into the control strategy.

Chapter 6

Conclusion and Perspectives

In this document, I have attempted to summarise the main results I have attained in the past seven years of research within the interdisciplinary and challenging domain of human mobility resilience. This contribution has been presented in terms of three main methodological axes: *(i)* mining of large and heterogeneous data, *(ii)* modelling and analysis of complex and large networks, and *(iii)* development of real-time big data platforms and algorithms.

These axes have been explored in the context of four different aspects of resilience: *(i)* learning from the past, *(ii)* anticipating what to expect, *(iii)* monitoring the critical properties of the system, and *(iv)* responding with the help of proper decision-making tools.

More specifically, the contribution that I consider the most relevant among the different results achieved are the following:

- We demonstrated the possibilities offered by passively collected mobile phone data, and, particularly, Network Signalling Data, to accurately estimate travel demand and traffic performance indicators, as well as precisely reconstruct mobility trajectories in urban environments.
- We have developed state-of-the-art approaches to k-anonymise trips reconstructed from mobile phone data, with minimal loss of information and reduced computation time, paving the way to the possibility of making the resulting datasets available to the community in an open, privacy-aware aggregate format (dynamic Origin-Destination matrices).
- We have combined solutions from two different but increasingly close fields - data science and complex network theory - to identify metrics and approaches that can help perform real-time monitoring of transportation networks and anticipate the occurrence of disruptive events.
- We have combined historical data analysis with numerical programming to identify resilient and sustainable on-demand flexible mobility services (*i.e.,* Park-and-Ride systems), which can complement the existing transport offer with a solution that can withstand a large set of likely-occurring disruptive scenarios.
- We have leveraged distributed computing, cloud and big data technologies to build the foundation of a novel extensible platform for real-time monitoring and control of large-scale networks, and, more in general, complex and interdependent urban systems.
- We have developed a state-of-the-art algorithm for the fast computation of Betweenness Centrality, a fundamental metric for discovering opinion leaders and vulnerable components of different kinds of networks.

Once more, I am grateful to all the researchers I have had the opportunity to collaborate with over the past years, as their invaluable contributions deserve a substantial portion of credit for these achievements.

In the rest of this chapter, I will highlight some ongoing works, presented as short-term perspectives (Sec. 6.1), that will further develop the main findings described above. The planned short-term research has fairly well-identified objectives and methodological bases, and I, therefore, expect conclusive results in the next few months and years. In section 6.2, I will instead outline some broader scientific directions that are likely to drive my research in the longer term, and for which the boundaries and precise methodologies are inevitably less clear. Nevertheless, I will try to identify more specific research themes or questions to help implement my research plan and facilitate the achievement of concrete results in the long term.

6.1 Short-term Perspectives

In the short term, my primary objective is to finalise and extend the work conducted within various research projects I have recently been or am still involved in. This section provides an overview of this ongoing work, capturing its current state and outlining my immediate future directions. These activities revolve around three main topics: *(i)* travel demand analysis and modelling with a special focus on multi-modal demand variations in the presence of disruptive events; *(ii)* accurate short-term traffic forecasting, in multi-modal settings and with respect to atypical events; and *(iii)* synthetic generation of presence and multi-modal mobility data. It is worth underlining that these themes all relate to some extent to enhancing the resilience of multi-modal transport networks, as will be further detailed below.

The first and second research subjects are strongly interrelated, as they globally aim to develop accurate models of multi-modal transport network dynamics by advancing the understanding of travel demand changes during non-recurrent events with respect to recurrent ones. These advancements will ultimately contribute to the state-of-the-art in the field of mobility management during atypical events. For the first topic, we are leveraging unsupervised machine learning solutions to identify and characterise travel demand variations in the presence of disrupted mobility episodes by mining heterogeneous data sources covering large periods of observations. Concerning the second one, we are exploring supervised deep learning frameworks to enhance the quality of short-term traffic prediction. Specifically, we are focusing on graph neural networks and transformer-based architectures, which represent the state-of-the-art in the field. The initial results of these studies are reported in the two subsections of Sec. 6.1.1.

The third research subject focuses on the generation of synthetic data by employing statistical inference techniques to fuse multiple available sources of real data, such as mobile telephony data, population statistics, survey data, and OD matrices. The goal of this research is to obtain rich and reliable synthetic population, travel demand and transport-mode choice data that closely mirror real-world conditions, as reported in the input data. These synthetic dataset can have useful applications in terms of resilience, enabling researchers to conduct in-depth and realistic vulnerability analyses and experiments without compromising user privacy. Ultimately, this research shall contribute to developing more robust and sustainable transportation design and management strategies. Our current approach, further detailed in Sec. 6.1.2, not only has the potential to enhance the accuracy of simulation tools but also ensures the privacy and anonymity of individuals represented in the generated data.

Parts of the following text are drawn from papers that are currently being written or under review and might be published as conference proceedings or journal papers between the writing of this chapter and the public defence of my HDR. Whenever possible, references to pre-print or preliminary versions are provided within the text.

6.1.1 Short-term Multi-modal Traffic Forecasting: looking for the atypical

Forecasting traffic-related variables, such as flow, speed or travel demand, plays a critical role in traffic monitoring and control as it allows for anticipating changes and proactively developing mitigation strategies to handle potential disruptions and ensure resilient mobility. In our ongoing research, we aim to tackle the challenge of understanding, modelling and predicting fluctuations in traffic variables during highly dynamic and non-recurrent situations, such as peak hours, holidays, and special or disruptive events. We believe that this topic represents an under-explored area of research that should occupy a central position at the intersection of traffic modelling and transport resilience engineering.

To achieve this objective, we are currently investigating the dynamics of multi-modal demand during atypical events as well as the integration into deep-learning forecasting architectures of multi-modal travel demand data and context-specific information, such as weather conditions, planned road closures, accidents, and public events. Our aim is to develop innovative machine learning solutions capable of accurately forecasting[1] multi-modal traffic variations in both regular and atypical traffic scenarios. Ultimately, such prediction tools can be integrated into scenario analyses and decision-making solutions aimed to optimise the response of the transport network to stress situations. Existing literature on this topic is limited [251]. Traditional forecasting methods are typically designed for stable time frames and specific geographical areas, neglecting rarer and more uncertain situations. Extreme examples of such dynamic conditions that we would like to address include the COVID-19 pandemic, population protests, and exposure to climate-change-related events like heavy rainfall, floods and heatwaves.

The preliminary outcomes of our research on the topic are reported below and have been mostly obtained in the context of the ANR PRCE MOBITIC project and the PhD theses of Ali Benam and Romain Rochas. These results represent our first steps towards answering the following research questions: *(i)* Can we automatically identify and model categories of irregular travel demand behaviours and use them to further improve short-term multi-modal traffic prediction? *(ii)* Which deep-learning frameworks are most suitable for short-term traffic prediction? *(iii)* Are graph neural network solutions appropriate for predicting multi-modal travel demand? *(iv)* What level of performance can be achieved in perturbed and unstable mobility scenarios?

Moreover, the study of travel volume variations and modal shifts due to abnormal events is at the core of the research that will be carried out within the scope of Paul Denailly's postdoctoral fellowship. This research is funded by the ANR MOBITIC project and supervised in collaboration with the GRETTIA lab at the University Gustave Eiffel, focusing on the identification and understanding, via data fusion techniques, of the impacts on mobility induced by the COVID-19 pandemic.

Multi-source data-driven identification and classification of multi-modal travel demand variations during atypical events

We have developed a preliminary methodology to automatically detect and analyse multi-modal travel demand variations following disruptive events. Currently, the methodology can be applied to perform anomaly detection at a network-wide scale and is partially inspired by the mobile-phone signature-based approach described in Sec. 2.1.2.

Previous research [252, 253] has examined the passenger responses to disruptions in transport systems using surveys and mobile or public transport ticketing data to analyse the

[1] We consider traffic forecasting as related to predicting, *e.g.,* OD pairs travel demand or average speed at the level of road sections, zones or the entire transport network, depending on the specific case study. We refer to short-term prediction with respect to prediction horizons in the order of minutes or hours, by relying on historical time series of several hours, days or a week.

demand dynamics and shifts between two specific modes during certain events. However, current literature lacks an integrated, automated approach that goes beyond single-kind disruption events and accounts for multiple urban transport modes. Our work provides a general data-driven methodology for the identification and categorisation of complex multi-modal demand dynamics during disruptive events.

The developed methodological framework comprises four main steps, as detailed in Figure 6.1, and has been evaluated in a case study related to the multi-modal transport demand for the city of Lyon, France. A brief summary is provided below, while more details can be found in [254].

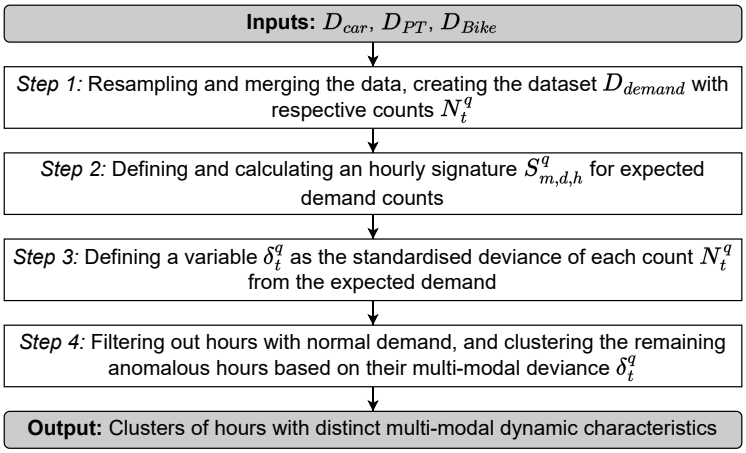


Figure 6.1: Methodology flowchart

In *step 1*, a compressed representation of the typical demand for each transport mode is constructed by unifying multi-source datasets. Traffic observations are aggregated hourly, per each mode. A unified dataset is thus produced with demand counts, denoted as $D_{\text{demand}} = \{N_t^q\}$. N_t^q represents the observed demand count for mode q on the generic 1-hour time slot t from the data availability period. In our case study, $q \in Q = \{\text{bus, tram, metro, bike, car}\}$.

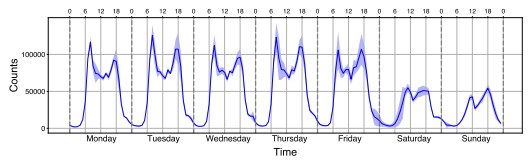
In *step 2*, the methodology defines a *rolling signature* $S^q_{(m,d,h)}$ that represents the expected demand for the generic mode q over a given reference period (*e.g.,* a given week m, on the generic hour h of day of the week d). The definition of signature is based on a compressed representation combining the hourly mean (μ) and standard deviation (λ) counts computed from past and future demand observations. The temporal support (W_m^K, d, h) used for the aggregation includes a subset of weeks ($W_m^K \subset M$), with K indicating the number of past and future weeks (with respect to week m) used for the aggregation, selected weekdays ($d \in \Delta$), and selected hours of the day ($h \in H$).

$$S^q_{(m,d,h)} = \mu^q_{(W_m^K,d,h)} \pm \lambda^q_{(W_m^K,d,h)} \tag{6.1}$$

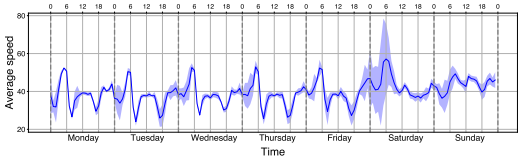
An example of a signature for each transport mode is reported in Fig. 6.2.

In *step 3*, the continuous variable δ_t^q is computed as formulated in Equation 6.2. δ_t^q represents the standardised deviation of the observed demand from the expected demand for a particular mode at a given hour. By pivoting the data on the deviance variable δ_t^q, we produce a dataset D_{deviance} with transport modes for columns, hours for rows, and standardised deviance of expected demand for values.

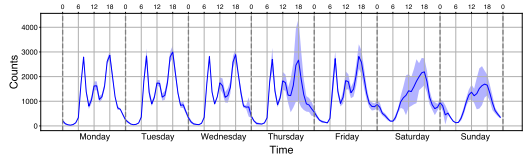
$$\delta_t^q = \frac{N_t^q - \mu^q_{(W_m^K,d,h)}}{\sigma^q_{(W_m^K,d,h)}} \tag{6.2}$$



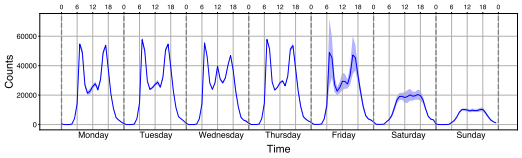
(a) The signature for the number of car observations



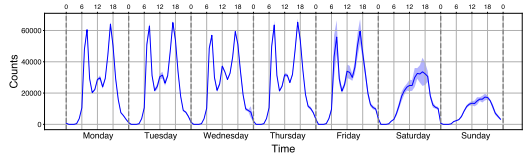
(b) The signature for the average recorded speed of cars



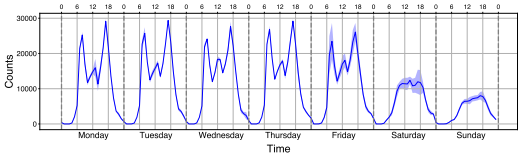
(c) The signature for the number of bike retractions



(d) The signature for the number of bus validations



(e) The signature for the number of metro validations



(f) The signature for the number of tram validations

Figure 6.2: The weekly signature of counts and observations for a week in March 2019

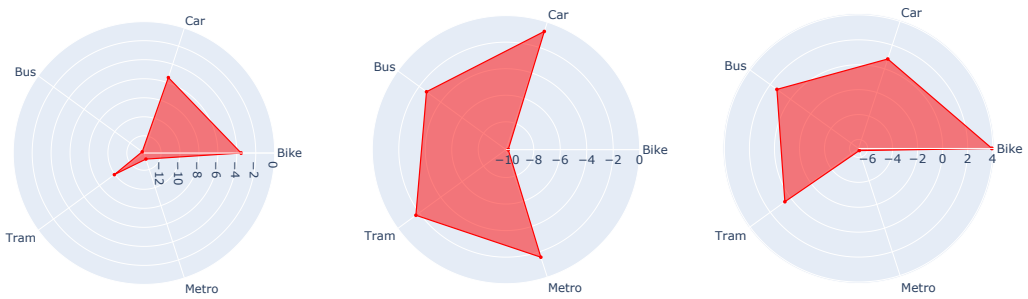
In *step 4*, as we are solely looking through anomalous hours for their categorisation, we omit hours in which demand in all modes falls inside the expected signature in a new dataset denoted as D_{anomaly}. This dataset is produced by filtering out regular hours; thus, it only includes hours where at least one mode's absolute value of deviance $|\delta_t^q|$ is larger than the range amplitude of the signature $\lambda^q_{(W_m^K,d,h)}$. We then apply clustering to this dataset to identify groups of time slots corresponding to categories of anomalies regarding multi-modal demand dynamics.

For the evaluation of the methodology, we focused on road traffic GPS observations, public transport ticket validations (tap-in for trams, metro, and bus lines) and shared-bike trips in the greater Lyon area, France, between 2019 and 2020. These sources of data have been simplistically leveraged as proxies of travel demand for the corresponding mode (*i.e.,* motorised vehicles, transit and cycling).

For each identified cluster, we selected all hours belonging to the cluster to study how that cluster manifested a particular multi-modal dynamic. We averaged the deviance δ_t^q for each mode in the given cluster and produced a multi-modal demand profile for each cluster. We then used radar plots to illustrate these multi-modal profiles. Additionally, we labelled the dates and the hours of anomalies belonging to each cluster to explore their temporal correspondence to traceable events.

Fig. 6.3 shows three clusters of anomalous data using radar plots to illustrate the average deviance of each mode in each cluster. Cluster 1 indicates lower public transport demand on national holidays, Cluster 2 shows decreased shared-bike use during periods of higher precipitation, and Cluster 3 represents metro service disruptions with passengers shifting to other modes. The findings were validated using news and weather data, supporting the explainability of the multi-modal dynamics within the clusters.

The results presented in this section instil confidence in the potential to develop a comprehensive framework for detecting and analysing the spatio-temporal dynamics of the multi-modal travel demand. Ultimately, our goal is to inspect the retrieved anomalous dynamics and infer a general model of modal shifts in the presence of the associated non-recurrent



(a) Cluster 1. (b) Cluster 2. (c) Cluster 3.

Figure 6.3: Average deviance of each mode per cluster (three largest identified clusters).

events, at various spatio-temporal granularities. While the current approach focuses on network-wide resolution, we are exploring the feasibility of defining reliable signatures at the level of public transport lines and geographical areas (*e.g.,* TAZ, IRIS zones, etc.) with promising initial results. By defining spatially-anchored multi-modal signatures, it should become possible to explore smaller-scale anomalies in travel demand, such as those due to localised events and disruptions. This approach is expected to facilitate the study of demand variation propagation in both space and time, thus offering the opportunity to identify areas, transport modes, or lines that are most vulnerable, *i.e.,* susceptible to disruptions, and with less accessibility, *i.e.,* with reduced possibility of modal shifts. These areas should be the target of transport policy aimed at reinforcing the mobility offer of the city by preserving equal access to transportation means.

Our short-term plans also involve integrating the signature-based approach with deep-learning forecasting models, which will be presented in the next subsection. By training such models with historical data related to multi-modal travel demand variations and contextual details of the corresponding anomalies, an ambitious objective would be to create forecasting tools capable of anticipating the increase or decrease in travel demand for one or multiple modes when specific disruptions are likely to occur or known to have recently occurred in an area of the city or network-wide. This research shall extend the state-of-the-art in deep learning-based traffic forecasting solutions for transport resilience, by enabling accurate predictions in both recurrent and non-recurrent scenarios and proactive decision-making in response to disruptive events.

Travel demand prediction during atypical events

The importance of traffic forecasting applies to all modes of transportation, including public transport and soft modes such as private and shared cycling. By accurately knowing such demand and anticipating its variations over time, transport operators can plan, develop and adapt the city mobility infrastructure to accommodate varying needs, promoting active and sustainable transportation, and reducing dependence on cars. Furthermore, incorporating context information into traffic prediction models is crucial for improving accuracy and enhancing the usefulness of forecasts. Factors like weather conditions, such as rain, snow, or extreme temperatures, can significantly impact travel demand and traffic patterns. For example, during inclement weather, travellers may choose to abandon active modes and switch to public transportation or private cars, leading to shifts in demand and traffic congestion. By integrating meteorological data into forecasting models, transportation authorities can anticipate these changes and implement appropriate measures, such as adjusting public transportation schedules or providing real-time information to connected vehicles. Special events, such as concerts, sports games, or festivals, can also have a significant impact on

traffic patterns and demand. Predicting the influence of these events allows transportation authorities to prepare for increased demand, provide additional transportation services, or implement temporary traffic management measures to minimise disruptions. In a similar way, incorporating real-time data on accidents, road closures, or construction activities can improve forecasting accuracy, allowing authorities to inform travellers about alternative routes to mitigate congestion.

With these objectives in mind, we are currently investigating the performance of state-of-the-art Graph Neural Networks (GNN) and Graph Convolutional Neural Networks (GCNN)-based solutions for travel demand prediction. As a preliminary step, we have focused on single-mode OD travel demand prediction for shared bikes, with the goal to highlight the limitations of such solutions when it comes to anticipate variations of demand in the presence of disruptive events. Bike-sharing systems serve as connection on the multi-modal transport network [255], allowing for reduced travel times, cost-effectiveness, and a smaller spatial and environmental footprint compared to traditional modes of transportation. While demand forecasting in this domain has historically relied on statistical and machine learning techniques, the advent of deep learning, and, more recently, GNNs, has ushered in a new era of forecasting algorithms, promising improved accuracy. Various models exist for bike-sharing demand prediction, ranging from station level forecast [256], to cluster-based prediction [257], and many other variations deeply detailed in surveys such as [258]. However, only a limited number of studies have focused on OD travel demand prediction [259]. OD prediction poses unique challenges, as the final destinations are usually unknown, and OD matrices generated are sparse [260]. Moreover, forecasting demand in the case of atypical phenomena, such as non-recurrent events, remains a challenge [261]. Bike-sharing demands exhibit temporal and calendar dependencies tied to factors such as weekdays, public holidays, workdays, and school holidays [262]. Furthermore, they are influenced by weather conditions such as wind, humidity, temperature, and, particularly, rainfall [262].

We have chosen to investigate a state-of-the-art prediction algorithm, the Spatio-Temporal Encoder-Decoder Residual Multi-Graph Convolutional network (ST-ED-RMGC), recently proposed by Ke *et al.* [260] to perform bike-sharing OD demand forecast. ST-ED-RMGC consists in an encoder-decoder based model, where the encoder is composed by a temporal module which takes into account the temporal dependencies of all OD pairs, and a spatial module with several Residual Multi-graph Convolutional networks (RMGCs), which take several adjacency matrix and the graph OD demand as inputs. The RMGC combines a residual module with a multi-graph convolution to capture the spatial correlation between OD pairs. The residual module is introduced to tackle the issue of gradient explosion/gradient vanishing in complex deep networks. The multi-graph convolution applies graph convolution on stacked weighted adjacency matrices.

By focusing on the prediction of bike-sharing trip count between IRIS zones of the city of Lyon, we have explored the possibility of incorporating contextual information such as time, weather conditions, and multi-modal data (*e.g.,* road car flow data) in order to improve the forecasting accuracy of the original ST-ED-RMGC algorithm. The details of the proposed enhanced ST-ED-RMGC architecture can be found in [263, 264].

Both the extended and the original ST-ED-RMGC algorithms have been trained on a dataset including hourly trip counts from January 8th, 2019 to November 9th, 2019. We considered 2500 ODs of the selected territory covering the cities of Lyon and Villeurbanne, France. For performance evaluation, we have considered a test dataset covering the 7 a.m. and 9 p.m. period for all days between December 1$^{\text{th}}$, 2019, and March 10$^{\text{th}}$, 2020. Alongside this test dataset, named *global scenario* in the following, we also identified specific *weather-related scenarios* to assess the accuracy of bike-sharing demand prediction under degraded weather conditions. All these scenarios were employed to make a comparative evaluation between the original ST-ED-RMGC, used as a baseline, and our enhanced model, which incorporates contextual data.

To construct the weather-related scenarios, we selected subsets from the global scenario

based on two factors computed from the available weather-related historical time series: the observed daily cumulative rainfall level (dcr) and the hourly amount of rain (hr). Since the choice of bike sharing as a mode of travel can be greatly influenced by ongoing, expected or recently concluded rainfall episodes, we considered in our scenarios: peak hours with rain (denoted as $hr > 0$), peak hours without rain (denoted as $hr = 0$), and peak hours with a significant level of daily cumulative rainfall. The metrics used to evaluate the models were the Mean Squared Error (MSE) and the MAPE. With no surprise, in scenarios without rain ($hr = 0, dcr = 0$) and in the global scenario, the variations of the ST-ED-RMGC model with weather features (both historical and forecast) exhibit only marginal improvement compared to the baseline ST-ED-RMGC model. However, in rainfall scenarios ($hr > 0$), all models incorporating weather information outperform the baseline, showing up to 20% and 27% improvements in MSE and MAPE, respectively. Notably, rainfall duration emerged as a relevant predictor of bike-sharing demand during degraded weather conditions. The initial findings of this study emphasise the significance of the considered contextual data, particularly weather and calendar-related information, in forecasting bike-sharing demand, especially during adverse weather conditions.

Despite such initial promising results, further investigation is needed, particularly regarding non-recurrent events beyond weather-related occurrences. First of all, this study serves as a stepping stone towards extending the proposed approach to encompass multi-modal travel demand prediction. We are currently working on modifying the original architecture for incorporating data on public transportation usage, car flow, and speed data. This integration aims to develop a versatile framework that could capture inter-modal travel demand relationships jointly with contextual data to perform accurate multi-modal travel forecasting. Although the literature on this subject is relatively limited, the work from Liang *et al.* [265] presents an interesting research direction. The authors of this paper combine multi-relational spatio-temporal GNN with attention-based mechanisms to capture the complex spatio-temporal patterns across multiple modes and accurately predict multi-modal demand.

Another notable challenge that we aim to address in the short-term is the scarcity of atypical events in the available datasets. To address this issue, an intriguing research direction involves exploring transfer-learning-based approaches to test the trained model on diverse traffic-related datasets from various cities and transport modes, thereby expanding its applicability.

As another research direction, we are attempting to leverage additional data sources, particularly mobile phone app consumption data, to develop innovative traffic forecasting solutions capable to better capture dynamic contextual information on both human presence and human mobility to anticipate traffic variations. Mobile phone app consumption data offer valuable insights into the activities individuals perform at the specific locations where they use their mobile phones, such as browsing the web, utilising navigation apps, streaming audio/video content, or engaging with social media platforms. Sudden fluctuations in the overall usage of a particular service could serve as an indicator of the initiation of an event, which, in turn, may result in subsequent traffic state or demand changes. For instance, a sudden surge in social network usage might be indicative of individuals witnessing an attack or an extreme event, which could potentially lead to a mass evacuation in the near future. By incorporating such real-time mobile app consumption data into our analysis, we aim to enhance our ability to detect and respond to emerging mobility patterns, ultimately enabling more effective management of transportation systems in response to unforeseen events.

Finally, a more ambitious research direction will focus on integrating traditional traffic flow models with neural-network-based forecasting algorithms towards increased accuracy, interpretability and adaptability to unseen traffic scenarios. This approach follows the principles of physics-guided machine learning, a methodology recently investigated in traffic forecasting with promising results [266]. Traffic flow models typically encapsulate conservation laws that are similar to those observed in fluid dynamics. The incorporation of these physical laws into machine learning frameworks serves to enhance not only predictive accuracy

but also the model's level of interpretability. This effectively mitigates a significant limitation often associated with purely data-driven methods. When calibrated with empirically derived travel demand data and substantiated by realistic mode choice models, these physics-informed machine learning models should be expected to faithfully reproduce complex traffic scenarios. Such models can be, therefore, leveraged to swiftly detect discrepancies between real-time data and projected traffic patterns. This capability should enable more informed and timely anomaly detection, contributing to more resilient traffic management strategies. Concurrently, it may be worthwhile to investigate emergent learning frameworks such as Liquid Neural Networks (LNN), initially introduced in [267]. LNN constitute a category of brain-inspired, continuous-time neural-network models that demonstrate capability to capture causal relationships and adaptability under changing conditions, as evidenced in [268], where they were successfully deployed for robust flight navigation using autonomous drones in unpredictable and out-of-distribution scenarios. The integration of biologically inspired priors into LNN results in flexible, robust, and interpretable forecasting solutions that can be adapted to new environments and conditions. These attributes make LNN particularly well-suited for developing versatile learning frameworks capable of performing forecasting across a diverse array of rapidly evolving and unobserved traffic scenarios.

6.1.2 Synthetic Population Generation via Multi-Source Data

Agent-based models in transportation traditionally rely on synthetic travel demand, which involves creating a synthetic population of households and individuals with socio-demographic attributes, along with daily activity patterns in both time and space. This approach allows for studying traffic evolution and evaluating the impact of specific transport supply and travel demand variations. However, for the generated simulations to be realistic and meaningful, it is crucial that the synthetic population accurately represents the actual population and its multi-modal travel practices.

Creating a realistic synthetic population of agents is a challenging task, traditionally involving multiple data sources that are processed according to the following steps [269]: *(i)* ensuring that the marginal distributions of the socio-economic features of the actual population, such as gender, age, and occupation, known from sources like census data, match those of the synthetic population; *(ii)* associating daily activity patterns to the synthesised population by statistically matching the trip chains derived from a second smaller-scale input population sample, such as regional household travel surveys.

A critical limitation of current synthetic population generation approaches is that they traditionally match activity chains without synthesising realistic location information for activities due to the difficulty in identifying reliable and high-resolution spatial data sources for agent locations. While the home location is generally known via census data, it is rarer that work/study and secondary activity locations are actually inferred from real data. The main methods used to generate location activities include gravity models for commuting patterns [270], discrete destination choice models for workplace assignments [271], and applying space–time constraints for the selection of discretionary activity locations [272]. However, while these approaches can produce reasonable trips in some circumstances, the reliability of the final synthetic travel demand remains limited compared to real-world-data matching possibilities.

We are currently working on extending the state-of-the-art pipeline proposed in [269] with a novel data-driven approach for the location of activities, leveraging anonymised mobile phone data from the telecommunications provider Orange. As complete trajectories can be computationally heavy to process, and feature a significant privacy risk, our idea consists in using the time-dependent OD matrix from mobile-phone NSD reconstructed trajectories [273], obtained with the approaches described in Sec. 3.2 and Sec. 3.4. OD matrices are lighter and more manageable than whole trajectory datasets. This makes them more readily available, easier to process and safer regarding the privacy of transportation users. More-

over, they are built based on high-penetration and high-spatial-resolution data on human mobility 3.2, which makes them an ideal candidate for realistically reconstructing activity locations.

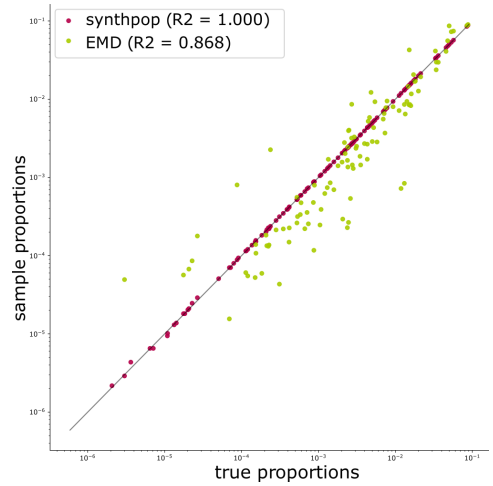
The main originality of the proposed approach lies in interpreting the OD matrix as a transition probability between two locations, which is used to sample the location chain. Additionally, the total number of trips in the OD matrix for each time of the day is utilised as a target to match the population agendas, achieved through hierarchical re-scaling for individuals and trips.

Currently, multiple alternative approaches are being investigated to draw location information associated with a whole activity chain of an agent from the OD-based transition probability matrix. While census data are used to determine a pre-defined home location for each agent, unlike state-of-the-art methods, we do not rely upon sources other than the OD matrix to pre-define work, study, or secondary activity places. This decision is justified by the fact that 45% of trips in the activity chains of the population are commute trips, while the mobile data contain the distribution of the sum of commute and non-commute trips. Using them only for the latter purpose would be invalid.

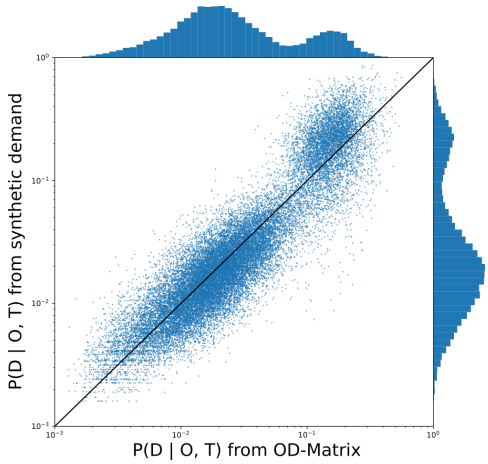
Among the most promising solutions to associate location information with each agent's activity, an approach based on Markov Chain Monte Carlo (MCMC) sampling is being considered, integrated as an additional step to the base pipeline defined in [269]. Our MCMC-based sampling process begins by fixing the agent's home location from census data and ensuring that all work-related activities of any single agent are drawn only once from the OD-based transition probability matrix, remaining the same for all activity chains containing them, assuming each agent has only one work or study place. The MCMC sampling then proceeds by assigning known values to all pre-defined states (*e.g.,* the agent's known home location at the beginning and end of their activity chain) and random values to all other states. Subsequently, a new value is successively drawn for each state with dependency on the other states considered known. Through iterative sampling of the unknown locations a large number of times, MCMC converges to the stationary distribution within an acceptable error.

The proposed methods have been evaluated for an area of 25,600 meters wide centred around Lyon, divided into 515 distinct zones, representing municipalities or spatial units (IRIS) common to many statistical analyses in France. We used the census performed by the French statistics institute INSEE to generate the synthetic population of travellers. Each of the 487,628 rows for our study zone features socioeconomic variables and a non-integer scaling coefficient for a total of 1,366,072 persons. The activity chains were taken, without location information, from the household travel survey (named Enquête Ménage Déplacements (EMD)) performed by the French agency for urban planning [274], which details the complete agendas of 25,203 persons in the Lyon region. Finally, we used OD-matrices from TRANSIT spatio-temporally describing the movement of residents of the region of Lyon. These data cover three months, from 19$^{\text{th}}$ March 2019 to 18$^{\text{th}}$ June 2019 and contain information about 49,681,883 distinct flows. As the scope of the other data sources is restricted to working days, we removed weekends and main holidays in France.

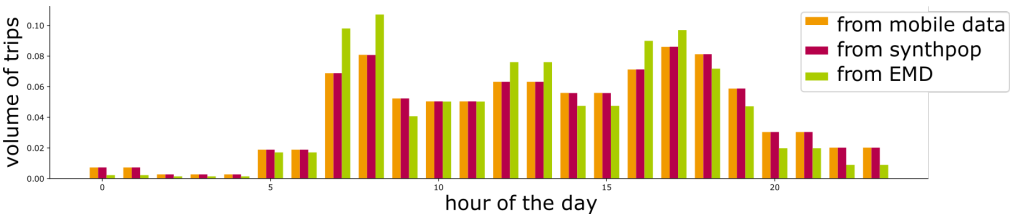
The current validation of the generated synthetic population primarily relies on verifying the agreement between the distributions from the input data (*i.e.,* census data, mobile phone OD matrix, and household travel surveys) to those derived from the synthetic population. Naturally, the algorithms aim to match the distributions derived from these sources, which should ideally correspond closely to those related to the synthetic population. However, discrepancies in the data sources pose challenges in effectively and jointly matching them. Among other limitations of the data sources, in particular, the socio-economic population sample is small and thus not fully representative of the entire population. In Fig. 6.4(a), we illustrate how the socioeconomic composition of our population matches the official census better than the EMD survey. Each red point represents the share of the population with a given socio-economic feature value, in the case of the actual one (on the x-axis) versus



(a) Matching of the socioeconomic marginals of our rescaled synthetic population (in red) compared to the HTS (in green).



(b) Distribution of trips taken during the day.



(c) Distribution of trips taken during the day.

Figure 6.4: Average deviance of each mode per cluster.

the synthetic one (on the y-axis). The same comparison is reported with green points for the EMD population. As the EMD represents only a small number of people, it is normal that the totals have a higher variance. Our synthetic population can then be seen as a version of the EMD that fully agrees with the census on the socioeconomic distribution of the population. Fig. 6.4(c) graphically presents the distribution of trips during a typical day as measured from mobile data (in orange), our synthetic data (in red), and the survey EMD (in green). In this case, we considered the observed volumes from the mobile data to be the ground truth. With no surprise, it can be noted that our population perfectly fits the ground truth[2]. On the contrary, the EMD seems to overestimate the volumes of the morning, noon, and evening peaks while underestimating the volumes during the rest of the day.

Finally, we compared the flows obtained with our approaches with the ones observed from the mobile data. A first natural way of comparing the flows is to compare the distributions of destinations $P(d|o,t)$. In Fig. 6.4(b), we illustrate how the probability of destinations given origin and hour fits the probability table derived from the OD matrix. Each point corresponds to a combination of destination, origin, and time of the day. Its x-coordinate is the probability $P(d|o,t)$ observed in the OD matrix, while its y-coordinate is the same probability as observed in the trips of our synthetic population. We observe a bi-modality in the ground truth, and a generally good fit of the synthetic population flows with the mobile phone OD matrix.

The reported results represent a preliminary step in developing the complete methodology, but further improvements and assessments must be considered. While it can be

[2]Here, we consider the mobile phone data trip volumes as ground truth.

expected that our synthetic demand data matches well the distribution $P(D|O,T)$, we have observed that the distribution of trips $P(O,D|T)$ is not the same as in the input OD matrix. This is because although we sample the destinations using $P(D|O,T)$ as a transition probability, we have no mechanism to ensure that the correct number of agents leave each origin O at each time step T. This problem can be addressed by decomposing each time slice of the transition matrix into a sum of transition matrices depending on O, T, and on the time step of the next trip of the agent. This new decomposition of the OD matrix in our future work amounts to adding explanatory variables to the mobile data. By carefully choosing such a decomposition, we can make sure the agents taking a trip before time step T are assigned to destinations such that the map of agents leaving for a trip at time step T corresponds to the map of origins of trips in the OD matrix. Additional future research directions include refining the structure of the OD transition matrix by distinguishing recurring commuting trips from non-recurring ones, as well as introducing additional constraints on trip distance and transport modes for more accurately matching location activities, by possibly including additional data sources in the pipeline. We also aim to perform validation using, as ground truth, external data sources that were not considered within the population generation pipeline. For instance, we plan to incorporate aggregate statistics on mobility in the considered region from different data providers to validate, *e.g.,* the distributions of trip distances, transport mode shares, and the joint distribution of activity locations based on socio-economic features. This validation will help us gain further confidence in the reliability of our methodology and outcomes.

Additionally, we aim to conduct multi-modal traffic simulations with in-house macroscopic simulators, properly calibrated with external data sources, including multi-modal travel demand data, loop detector flow data, FCD speed data, and ticket validation data. The resulting simulated traffic states will be compared to those obtained using a multi-agent multi-modal traffic simulator, such as MATSim [275], using our generated synthetic population as the main input to run simulations.

6.2 Longer-term Perspectives

In the long term, my primary research objective is to investigate and enhance the resilience characteristics of urban networks within a *multi-domain context* and according to a *holistic* vision. The ultimate goal is to contribute to developing an integrated, data-driven digital framework enabling *by-design* to assess and promote the joint resilience of diverse inter-related urban technical networks, such as those related to mobility, energy, telecommunications, and more.

These long-term perspectives find their reason in the urgency of addressing climate change, whose effects are especially critical in large metropolitan areas due to the presence of heterogeneous yet highly interdependent infrastructures [276, 277]. For example, electricity pylons collapsing from the consequences of extreme weather-related events can disrupt traffic by impeding vehicle evacuations, power outages can compromise mobility by affecting traffic lights and electrified transport means, and a lack of communication networks may hinder future connected and autonomous vehicle mobility.

Considering the fragility of these interconnected infrastructures, their susceptibility to climate change and the cascading effects of weather-related disruptions, it appears unavoidable to rethink the traditional *in-silo* approach to resilience modelling according to a new holistic approach aimed to investigate, model, monitor, and optimise *as a whole* the various networks constituting the urban infrastructure ecosystem, while prioritising energy efficiency, sustainability and user-privacy requirements.

The vision of resilience as a crucial urban commodity – involving the synergistic interaction of multiple stakeholders, resources and infrastructures – is becoming increasingly significant in each of the mentioned fields [278, 279, 280] and ties into the notion of Resilience-

as-a-Service (RaaS). The term was initially introduced in the context of high-performance computing [281] to provide cloud solutions with a highly available, distributed, and scalable fault-tolerant service. Very recently, the term has been introduced in the field of transportation [282], defining a conceptual and operational framework where the available resources of various service providers could be integrated to manage and improve over time the resilience of a system faced by hard-to-predict disruptions and extreme events.

Despite some isolated attempts to apply this concept in a multi-domain urban context [283, 284], significant research is still needed to implement and generalise this vision on a large scale, encompassing diverse infrastructural networks within a city. In particular, by considering again the parallel between the urban critical infrastructures and the cloud computing paradigm, I believe this vision should further emphasise the need for *targeted and energy-efficient data collection and processing*, *network-level sensing*, *continuous and explainable learning*, as well as *real-time network monitoring and control*. These functionalities should all be conceived as pivotal elements of an integrated urban network digital architecture for the realistic and large-scale implementation of RaaS.

To achieve this, three key macro-objectives are described in the following as the pillars of my future research.

The City Digital Twin

A key research objective involves the joint modelling of urban, multi-domain, and multi-modal networks as an integrated and autonomous digital system that mirrors physical reality, and which could be named the *city digital twin*. This objective aims to leverage real-time, multi-source data – spanning malfunctions, service usages, and contextual information from various service networks – as primary resources. These data would enable automated solutions to identify standard operational patterns, detect anomalies, and autonomously implement mitigation, adaptation, and emergency-response strategies. This vision is gaining more concrete ground for real-world implementation from emerging network infrastructures, including 6G telecommunication networks [285], IoT-enabled smart grids and micro-grids [286], fleets of autonomous connected vehicles for public transport [287], which are expected to integrate sensing and actuating mechanisms as standard features. The objective aligns closely with the expanding scope of *urban digital twins* [38, 288], a concept referring to the widespread adoption of digital twins technology across the urban ecosystem. This includes buildings, energy systems, vehicles, and telecommunication facilities and devices. A *city digital twin* shall push this vision further, thus functioning as an accurate digital replica of the physical urban environment *as a whole*. The city digital twin shall be endowed with advanced sensing capabilities, and autonomous interactions with the physical urban infrastructures, and provide self-diagnostic and self-repair functions that are aware of the inter-connections and inter-dependencies among the different technical components of the urban environment.

Coupled Vulnerability Assessment

A pivotal focus will be the identification and systematic analysis of critical interdependencies across a range of modelled networks. This effort aims to cultivate a modular and extensible platform specifically engineered for conducting coupled vulnerability assessments. The platform shall be capable of modelling, simulating, and quantifying the impacts of a diverse set of disruptive scenarios on various urban networks. A particular emphasis will be placed on understanding cascading effects, which result from failures in one network influencing others [289]. Central to this aim will be the design of the platform's core as middleware, crafted to facilitate seamless integration among a variety of simulation and modelling tools. This core will serve as the backbone for a unified framework wherein dynamic models for, *e.g.,* flood, communication, electrical, and traffic simulation, can coexist and dynamically

interact. Such interaction will enable these models to respond collaboratively to simulated stressors like heavy rainfall, heatwaves and other kinds of shocks. This integrated approach will provide holistic insights into the vulnerabilities inherent in the urban environment and allow for the assessment of the effectiveness of various resilience strategies.

Multi-domain Resilience Strategies

The last fundamental objective I would like to develop concerns the exploration of decision-making tools and strategies that promote the synergistic integration of various urban networks and services. The aim is to enhance the sustainability, robustness, adaptability, and mitigation capabilities of the urban system in the face of extreme events and network failures. For instance, optimal scheduling of mitigation and repair actions – such as reinforcing distribution poles or towers, pre-installing battery backup systems at signalised intersections, deploying emergency vehicles, and rerouting traffic to facilitate access to disrupted facilities – serves as a straightforward example aimed at bolstering the resilience of the integrated transport, electricity, and telecommunications systems. More ambitious initiatives could focus on the joint optimisation and emergency operation of the integrated urban networks. These can be enabled by the monitoring capabilities of the city digital replica and by the insights derived from the joint vulnerability assessments, in line with approaches very recently proposed in the literature [289]. For example, Electric and electrified Vehicles (EVs) can serve as emergency mobile resilience assets during prolonged power outages, supplying power to critical buildings like hospitals. Similarly, a fleet of EVs could power nearby facilities or even assist in restarting substation transformers following a natural disaster [290].

Energy efficiency, reduced carbon footprint and cybernetic resilience shall also be considered central requirements in this vision for a resilient and sustainable urban landscape. For instance, I plan to explore data-frugal processing frameworks, such as continual learning [291], which aim to optimise the capabilities of machine learning models. The objective here is to store key information from previous data observations, minimise the impact of catastrophic forgetting [292], and adapt swiftly to changing data distributions. Achieving this allows for a reduction in data footprint. Specifically, it reduces the computational and storage resources needed for data processing and analysis.

These perspectives and objectives will serve as the backbone for a European Research Council (ERC) consolidator project that I plan to submit in the coming years. Additionally, I am actively contributing to the development of a large-scale project on this topic with the Gustave Eiffel Foundation. My involvement includes participating in a dedicated working group on urban resilience that focuses on the study of the vulnerabilities of city technical networks, and the definition of proper adaptation strategies.

Bibliography

[1] L. Bevilacqua, A. Furno, V. S. Di Carlo, and E. Zimeo, "A tool for automatic generation of ws-bpel compositions from owl-s described services," in *2011 5th International Conference on Software, Knowledge Information, Industrial Management and Applications (SKIMA) Proceedings*. IEEE, 2011, pp. 1–8. 2

[2] A. Furno and E. Zimeo, "Context-aware composition of semantic web services," *Mobile Networks and Applications*, vol. 19, no. 2, pp. 235–248, 2014. 2

[3] ——, "Self-scaling cooperative discovery of service compositions in unstructured p2p networks," *Journal of Parallel and Distributed Computing*, vol. 74, no. 10, pp. 2994–3025, 2014. 2

[4] A. Furno, "Scalable service composition in autonomic computing," PhD dissertation, Univesity of Sannio, Department of Engineering, 2014. [Online]. Available: https://people.licit-lyon.eu/furno/documents/phd_thesis_angelo_furno.pdf 2

[5] P. Suppa and E. Zimeo, "A context-aware mashup recommender based on social networks data mining and user activities," in *2016 IEEE International Conference on Smart Computing (SMARTCOMP)*. IEEE, 2016, pp. 1–6. 2

[6] A. Furno, D. Naboulsi, R. Stanica, and M. Fiore, "Mobile demand profiling for cellular cognitive networking," *IEEE Transactions on Mobile Computing*, vol. 16, no. 3, pp. 772–786, 2016. 3, 26, 27

[7] D. Naboulsi, R. Stanica, and M. Fiore, "Classifying call profiles in large-scale mobile traffic datasets," in *IEEE INFOCOM 2014-IEEE conference on computer communications*. IEEE, 2014, pp. 1806–1814. 3, 26, 29

[8] D. Bega, M. Gramaglia, M. Fiore, A. Banchs, and X. Costa-Perez, "Deepcog: Cognitive network management in sliced 5g networks with deep learning," in *IEEE INFOCOM 2019-IEEE conference on computer communications*. IEEE, 2019, pp. 280–288. 3

[9] A. Furno, M. Fiore, R. Stanica, C. Ziemlicki, and Z. Smoreda, "A tale of ten cities: Characterizing signatures of mobile traffic in urban areas," *IEEE Transactions on Mobile Computing*, vol. 16, no. 10, pp. 2682–2696, 2016. 3, 10, 29, 31

[10] A. Furno, M. Fiore, and R. Stanica, "Joint spatial and temporal classification of mobile traffic demands," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9. 3, 10, 34

[11] C. Kennedy, S. Pincetl, and P. Bunje, "The study of urban metabolism and its applications to urban planning and design," *Environmental pollution*, vol. 159, no. 8-9, pp. 1965–1973, 2011. 3

[12] D. N. Bristow and C. A. Kennedy, "Urban metabolism and the energy stored in cities: Implications for resilience," *Journal of Industrial Ecology*, vol. 17, no. 5, pp. 656–667, 2013. 3

[13] E. Hollnagel, D. D. Woods, and N. Leveson, *Resilience engineering: Concepts and precepts*. Ashgate Publishing, Ltd., 2006. 5, 14, 16, 19

[14] E. Bellini and P. Nesi, "Exploiting smart technologies to build smart resilient cities," in *Routledge handbook of sustainable and resilient infrastructure*. Routledge, 2018, pp. 685–705. 5, 17

[15] V. Mahajan, N. Kuehnel, A. Intzevidou, G. Cantelmo, R. Moeckel, and C. Antoniou, "Data to the people: a review of public and proprietary data for transport models," *Transport reviews*, vol. 42, no. 4, pp. 415–440, 2022. 7

[16] C. Wang and D. B. Hess, "Role of urban big data in travel behavior research," *Transportation research record*, vol. 2675, no. 4, pp. 222–233, 2021. 7

[17] X. Kong, M. Li, K. Ma, K. Tian, M. Wang, Z. Ning, and F. Xia, "Big trajectory data: A survey of applications and services," *IEEE Access*, vol. 6, pp. 58 295–58 306, 2018. 7

[18] P. R. Stopher and S. P. Greaves, "Household travel surveys: Where are we going?" *Transportation Research Part A: Policy and Practice*, vol. 41, no. 5, pp. 367–381, 2007. 8, 35

[19] M. Lin and W.-J. Hsu, "Mining gps data for mobility patterns: A survey," *Pervasive and mobile computing*, vol. 12, pp. 1–16, 2014. 9

[20] O. Cats, "Identifying human mobility patterns using smart card data," *arXiv preprint arXiv:2208.05352*, 2022. 9

[21] S. Bührmann, "New seamless mobility services: Public bicycles," *NICHES Policy Notes*, 2007. 9

[22] S. A. Shaheen, S. Guzman, and H. Zhang, "Bikesharing in europe, the americas, and asia: past, present, and future," *Transportation research record*, vol. 2143, no. 1, pp. 159–167, 2010. 9

[23] J. Song, L. Zhang, Z. Qin, and M. A. Ramli, "Spatiotemporal evolving patterns of bike-share mobility networks and their associations with land-use conditions before and after the covid-19 outbreak," *Physica A: Statistical Mechanics and its Applications*, vol. 592, p. 126819, 2022. 9

[24] J. Zhou, Y. Guo, J. Sun, E. Yu, and R. Wang, "Review of bike-sharing system studies using bibliometrics method," *Journal of traffic and transportation engineering (English edition)*, vol. 9, no. 4, pp. 608–630, 2022. 9

[25] T. H. Silva, A. C. Viana, F. Benevenuto, L. Villas, J. Salles, A. Loureiro, and D. Quercia, "Urban computing leveraging location-based social network data: a survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 1, pp. 1–39, 2019. 10

[26] V. D. Blondel, A. Decuyper, and G. Krings, "A survey of results on mobile phone datasets analysis," *EPJ data science*, vol. 4, pp. 1–55, 2015. 10

[27] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 6 2008. [Online]. Available: http://www.nature.com/doifinder/10.1038/nature06958 10

[28] M. S. Iqbal, C. F. Choudhury, P. Wang, and M. C. González, "Development of origin–destination matrices using mobile phone call data," *Transportation Research Part C: Emerging Technologies*, vol. 40, pp. 63–74, 3 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0968090X14000059/ 10

[29] R. W. Douglass, D. A. Meyer, M. Ram, D. Rideout, and D. Song, "High resolution population estimates from telecommunications data," *EPJ Data Science*, vol. 4, no. 1, p. 4, 12 2015. [Online]. Available: http://www.epjdatascience.com/content/4/1/4 10

[30] A. Wesolowski, C. O. Buckee, K. Engø-Monsen, and C. J. E. Metcalf, "Connecting mobility to infectious diseases: the promise and limits of mobile phone data," *The Journal of infectious diseases*, vol. 214, no. suppl_4, pp. S414–S420, 2016. 10

[31] D. Naboulsi, M. Fiore, S. Ribot, and R. Stanica, "Large-scale mobile traffic analysis: a survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 124–161, 2015. 10

[32] Q. Xu, A. Gerber, Z. M. Mao, and J. Pang, "AccuLoc: Practical localization of performance measurements in 3G networks," in *MobiSys'11 - Compilation Proceedings of the 9th International Conference on Mobile Systems, Applications and Services and Co-located Workshops*. New York, New York, USA: ACM Press, 2011, pp. 183–195. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1999995.2000013 10

[33] G. Chen, A. C. Viana, M. Fiore, and C. Sarraute, "Complete trajectory reconstruction from sparse mobile phone data," *EPJ Data Science*, vol. 8, no. 1, pp. 1–24, 2019. 10

[34] G. Ranjan, H. Zang, Z.-L. Zhang, and J. Bolot, "Are call detail records biased for sampling human mobility?" *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 16, no. 3, pp. 33–44, 12 2012. [Online]. Available: https://dl.acm.org/doi/10.1145/2412096.2412101 10

[35] D. Bachir, V. Gauthier, M. E. Yacoubi, and G. Khodabandelou, "Using mobile phone data analysis for the estimation of daily urban dynamics," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 10 2017, pp. 626–632. [Online]. Available: http://ieeexplore.ieee.org/document/8317956/ 13

[36] H. R. Meredith, V. Andreani, H. R. Ma, A. J. Lopatkin, A. J. Lee, D. J. Anderson, G. Batt, and L. You, "Applying ecological resistance and resilience to dissect bacterial antibiotic responses," *Science Advances*, vol. 4, no. 12, p. eaau1873, 2018. 14

[37] T. Parsons, E. A. Shils, and N. J. Smelser, "The social system," in *Toward a General Theory of Action*. Routledge, 2017, pp. 190–233. 14

[38] X. Liu, D. Li, M. Ma, B. K. Szymanski, H. E. Stanley, and J. Gao, "Network resilience," *Physics Reports*, vol. 971, pp. 1–108, 2022. 14, 15, 19, 21, 22, 81, 157

[39] R. Billinton and R. N. Allan, *Reliability evaluation of engineering systems*. Springer, 1992, vol. 792. 14

[40] D. D. Woods, "Four concepts for resilience and the implications for the future of resilience engineering," *Reliability Engineering & System Safety*, vol. 141, pp. 5–9, 2015. 14, 15

[41] L. Fisher, "More than 70 ways to show resilience," *Nature*, vol. 518, no. 7537, pp. 35–35, 2015. 15

[42] M. Bruneau, S. E. Chang, R. T. Eguchi, G. C. Lee, T. D. O'Rourke, A. M. Reinhorn, M. Shinozuka, K. Tierney, W. A. Wallace, and D. Von Winterfeldt, "A framework to quantitatively assess and enhance the seismic resilience of communities," *Earthquake spectra*, vol. 19, no. 4, pp. 733–752, 2003. 15, 20

[43] D. Henry and J. E. Ramirez-Marquez, "Generic metrics and quantitative approaches for system resilience as a function of time," *Reliability Engineering & System Safety*, vol. 99, pp. 114–122, 2012. 15, 19, 20

[44] R. Steen and T. Aven, "A risk perspective suitable for resilience engineering," *Safety science*, vol. 49, no. 2, pp. 292–297, 2011. 15

[45] T. Aven, "Risk assessment and risk management: Review of recent advances on their foundation," *European Journal of Operational Research*, vol. 253, no. 1, pp. 1–13, 2016. 15

[46] L.-G. Mattsson and E. Jenelius, "Vulnerability and resilience of transport systems–a discussion of recent research," *Transportation research part A: policy and practice*, vol. 81, pp. 16–34, 2015. 15, 19, 20, 83

[47] J. S. Cañavera-Herrera, J. Tang, T. Nochta, and J. M. Schooling, "On the relation between 'resilience'and 'smartness': A critical review," *International Journal of Disaster Risk Reduction*, p. 102970, 2022. 17, 19

[48] K. Barker, J. H. Lambert, C. W. Zobel, A. H. Tapia, J. E. Ramirez-Marquez, L. Albert, C. D. Nicholson, and C. Caragea, "Defining resilience analytics for interdependent cyber-physical-social networks," *Sustainable and Resilient Infrastructure*, vol. 2, no. 2, pp. 59–67, 2017. 17

[49] M. Z. Serdar, M. Koç, and S. G. Al-Ghamdi, "Urban transportation networks resilience: indicators, disturbances, and assessment methods," *Sustainable Cities and Society*, vol. 76, p. 103452, 2022. 17, 19

[50] E. Henry, "Resilience modeling of urban multimodal transport networks," Ph.D. dissertation, Université de Lyon, 2021. 19

[51] Y. Cheng, E. A. Elsayed, and Z. Huang, "Systems resilience assessments: a review, framework and metrics," *International Journal of Production Research*, vol. 60, no. 2, pp. 595–622, 2022. 19

[52] B. Rathnayaka, C. Siriwardana, D. Robert, D. Amaratunga, and S. Setunge, "Improving the resilience of critical infrastructure: Evidence-based insights from a systematic literature review," *International Journal of Disaster Risk Reduction*, p. 103123, 2022. 19

[53] J. Douglas, "Physical vulnerability modelling in natural hazard risk assessment," *Natural Hazards and Earth System Sciences*, vol. 7, no. 2, pp. 283–288, 2007. 19

[54] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Physical review letters*, vol. 87, no. 19, p. 198701, 2001. 20

[55] P. Gauthier, A. Furno, and N.-E. El Faouzi, "Road network resilience: how to identify critical links subject to day-to-day disruptions," *Transportation research record*, vol. 2672, no. 1, pp. 54–65, 2018. 20, 83

[56] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *Journal of complex networks*, vol. 2, no. 3, pp. 203–271, 2014. 21, 69

[57] A. Aleta, S. Meloni, and Y. Moreno, "A multilayer perspective for the analysis of urban transportation systems," *Scientific reports*, vol. 7, no. 1, p. 44359, 2017. 21, 23

[58] J. Lin and Y. Ban, "Complex network topology of transportation systems," *Transport reviews*, vol. 33, no. 6, pp. 658–685, 2013. 21

[59] B. Jiang and C. Claramunt, "Topological analysis of urban street networks," *Environment and Planning B: Planning and design*, vol. 31, no. 1, pp. 151–162, 2004. 21

[60] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, no. 7291, pp. 1025–1028, 2010. 21

[61] Y. Duan and F. Lu, "Robustness of city road networks at different granularities," *Physica A: Statistical Mechanics and its Applications*, vol. 411, pp. 21–34, 2014. 21, 83

[62] E. Estrada, *The structure of complex networks: theory and applications*. Oxford University Press, 2012. 21

[63] J. Wang, "Resilience of self-organised and top-down planned cities—a case study on london and beijing street networks," *PloS one*, vol. 10, no. 12, p. e0141736, 2015. 21

[64] M. Akbarzadeh, S. Memarmontazerin, S. Derrible, and S. F. Salehi Reihani, "The role of travel demand and network centrality on the connectivity and resilience of an urban street system," *Transportation*, vol. 46, pp. 1127–1141, 2019. 21

[65] A. Nair and J. M. Vidal, "Supply network topology and robustness against disruptions–an investigation using multi-agent model," *International Journal of Production Research*, vol. 49, no. 5, pp. 1391–1404, 2011. 21

[66] E. Jenelius and L.-G. Mattsson, "Resilience of transport systems," in *Encyclopedia of Transportation*. Elsevier Amsterdam, Netherlands, 2020. 21, 81

[67] S. Dong, A. Mostafizi, H. Wang, J. Gao, and X. Li, "Measuring the topological robustness of transportation networks to disaster-induced failures: A percolation approach," *Journal of Infrastructure Systems*, vol. 26, no. 2, p. 04020009, 2020. 21

[68] J. Wu, Z. Gao, H. Sun, and H. Huang, "Urban transit system as a scale-free network," *Modern Physics Letters B*, vol. 18, no. 19n20, pp. 1043–1049, 2004. 21

[69] C. von Ferber, B. Berche, T. Holovatch, and Y. Holovatch, "A tale of two cities: Vulnerabilities of the london and paris transit networks," *Journal of Transportation Security*, vol. 5, pp. 199–216, 2012. 21

[70] Y. Meng, X. Tian, Z. Li, W. Zhou, Z. Zhou, and M. Zhong, "Comparison analysis on complex topological network models of urban rail transit: A case study of shenzhen metro in china," *Physica A: Statistical Mechanics and Its Applications*, vol. 559, p. 125031, 2020. 21

[71] C. Von Ferber, T. Holovatch, Y. Holovatch, and V. Palchykov, "Public transport networks: empirical analysis and modeling," *The European Physical Journal B*, vol. 68, pp. 261–275, 2009. 21

[72] C. Han and L. Liu, "Topological vulnerability of subway networks in china," in *2009 International Conference on Management and Service Science*. IEEE, 2009, pp. 1–4. 21

[73] M. E. O'Kelly, "Network hub structure and resilience," *Networks and Spatial Economics*, vol. 15, pp. 235–251, 2015. 22

[74] K. Adjetey-Bahun, J.-L. Planchet, B. Birregah, and E. Châtelet, "Railway transportation system's resilience: Integration of operating conditions into topological indicators," in *NOMS 2016-2016 IEEE/IFIP Network Operations and Management Symposium*. IEEE, 2016, pp. 1163–1168. 22

[75] Q.-C. Lu, "Modeling network resilience of rail transit under operational incidents," *Transportation Research Part A: Policy and Practice*, vol. 117, pp. 227–237, 2018. 22

[76] M. A. Taylor, "Travel through time: the story of research on travel time reliability," *Transportmetrica B: transport dynamics*, vol. 1, no. 3, pp. 174–194, 2013. 22

[77] R. D. Connors and D. P. Watling, "Assessing the demand vulnerability of equilibrium traffic networks via network aggregation," *Networks and Spatial Economics*, vol. 15, pp. 367–395, 2015. 22

[78] E. Jenelius, T. Petersen, and L.-G. Mattsson, "Importance and exposure in road network vulnerability analysis," *Transportation Research Part A: Policy and Practice*, vol. 40, no. 7, pp. 537–560, 2006. 22, 23, 84

[79] D. M. Scott, D. C. Novak, L. Aultman-Hall, and F. Guo, "Network robustness index: A new method for identifying critical links and evaluating the performance of transportation networks," *Journal of Transport Geography*, vol. 14, no. 3, pp. 215–227, 2006. 22, 23

[80] A. Nagurney and Q. Qiang, "Fragile networks: identifying vulnerabilities and synergies in an uncertain age," *International Transactions in Operational Research*, vol. 19, no. 1-2, pp. 123–160, 2012. 22

[81] E. Jenelius, "User inequity implications of road network vulnerability," *Journal of Transport and Land Use*, vol. 2, no. 3/4, pp. 57–73, 2010. 22

[82] J. L. Sullivan, D. C. Novak, L. Aultman-Hall, and D. M. Scott, "Identifying critical road segments and measuring system-wide robustness in transportation networks with isolating links: A link-based capacity-reduction approach," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 5, pp. 323–336, 2010. 22, 83, 85

[83] E. Jenelius and L.-G. Mattsson, "Road network vulnerability analysis of area-covering disruptions: A grid-based approach with case study," *Transportation research part A: policy and practice*, vol. 46, no. 5, pp. 746–760, 2012. 22, 91

[84] E. Mitsakis, A. Papanikolaou, G. Ayfadopoulou, J. Salanova, C. Doll, G. Giannopoulos, and C. Zerefos, "An integrated framework for linking climate change impacts to emergency adaptation strategies for transport networks," *European Transport Research Review*, vol. 6, no. 2, pp. 103–111, 2014. 22

[85] E. Jenelius, "Network structure and travel patterns: explaining the geographical disparities of road network vulnerability," *Journal of Transport Geography*, vol. 17, no. 3, pp. 234–244, 2009. 22, 23

[86] M. A. Taylor, S. V. Sekhar, and G. M. D'Este, "Application of accessibility based methods for vulnerability analysis of strategic road networks," *Networks and Spatial Economics*, vol. 6, pp. 267–291, 2006. 23

[87] M. A. Taylor *et al.*, "Remoteness and accessibility in the vulnerability analysis of regional road networks," *Transportation research part A: policy and practice*, vol. 46, no. 5, pp. 761–771, 2012. 23

[88] R. Nyberg and M. Johansson, "Indicators of road network vulnerability to storm-felled trees," *Natural hazards*, vol. 69, pp. 185–199, 2013. 23

[89] V. L. Knoop, M. Snelder, H. J. van Zuylen, and S. P. Hoogendoorn, "Link-level vulnerability indicators for real-world networks," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 5, pp. 843–854, 2012. 23

[90] X. He and H. X. Liu, "Modeling the day-to-day traffic evolution process after an unexpected network disruption," *Transportation Research Part B: Methodological*, vol. 46, no. 1, pp. 50–71, 2012. 23

[91] J. Yates and S. Sanjeevi, "A length-based, multiple-resource formulation for shortest path network interdiction problems in the transportation sector," *International Journal of Critical Infrastructure Protection*, vol. 6, no. 2, pp. 107–119, 2013. 23

[92] H. Ho, A. Sumalee, W. H. Lam, and W. Szeto, "A continuum modeling approach for network vulnerability analysis at regional scale," *Procedia-Social and Behavioral Sciences*, vol. 80, pp. 846–859, 2013. 23

[93] E. Rodríguez-Núñez and J. C. García-Palomares, "Measuring the vulnerability of public transport networks," *Journal of transport geography*, vol. 35, pp. 50–63, 2014. 23

[94] O. Cats and E. Jenelius, "Dynamic vulnerability analysis of public transport networks: mitigation effects of real-time information," *Networks and Spatial Economics*, vol. 14, pp. 435–463, 2014. 23

[95] ——, "Planning for the unexpected: The value of reserve capacity for public transport network robustness," *Transportation Research Part A: Policy and Practice*, vol. 81, pp. 47–61, 2015. 23

[96] S. K. Peterson and R. L. Church, "A framework for modeling rail transport vulnerability," *Growth and Change*, vol. 39, no. 4, pp. 617–641, 2008. 23

[97] L. Hong, M. Ouyang, S. Peeta, X. He, and Y. Yan, "Vulnerability assessment and mitigation for the chinese railway system under floods," *Reliability Engineering & System Safety*, vol. 137, pp. 58–68, 2015. 23

[98] M. Janić, "Modelling the resilience, friability and costs of an air transport network affected by a large-scale disruptive event," *Transportation Research Part A: Policy and Practice*, vol. 81, pp. 77–92, 2015. 23

[99] D. Yin, W. Huang, B. Shuai, H. Liu, and Y. Zhang, "Structural characteristics analysis and cascading failure impact analysis of urban rail transit network: From the perspective of multi-layer network," *Reliability Engineering & System Safety*, vol. 218, p. 108161, 2022. 23

[100] L. Bellocchi, V. Latora, and N. Geroliminis, "Dynamical efficiency for multimodal time-varying transportation networks," *Scientific reports*, vol. 11, no. 1, pp. 1–14, 2021. 24

[101] "Ericsson mobility report," Nov 2022. [Online]. Available: https://www.ericsson.com/4ae28d/assets/local/reports-papers/mobility-report/documents/2022/ericsson-mobility-report-november-2022.pdf 25

[102] Key technological challenges of the EC H2020 5G Infrastructure PPP. [Online]. Available: http://5g-ppp.eu/ 25

[103] V. D. Blondel, M. Esch, C. Chan, F. Clerot, P. Deville, E. Huens, F. Morlot, Z. Smoreda, and C. Ziemlicki, "Data for development: the d4d challenge on mobile phone data," 2012. [Online]. Available: https://arxiv.org/abs/1210.0137 26, 77

[104] Y.-A. de Montjoye, Z. Smoreda, R. Trinquart, C. Ziemlicki, and V. D. Blondel, "D4d-senegal: The second mobile phone data for development challenge," *ArXiv*, vol. abs/1407.4885, 2014. 26, 77

[105] T. Italia, "Telecom italia big data challenge," *URL https://dandelion.eu/datamine/open-big-data*, 2015. 26

[106] A. Ceselli, M. Fiore, A. Furno, M. Premoli, S. Secci, and R. Stanica, "Prescriptive analytics for mec orchestration," in *2018 IFIP Networking Conference (IFIP Networking) and Workshops*. IEEE, 2018, pp. 1–9. 28

[107] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, "Understanding traffic dynamics in cellular data networks," in *2011 Proceedings IEEE INFOCOM*. IEEE, 2011, pp. 882–890. 29

[108] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, S. Venkataraman, and J. Wang, "A first look at cellular network performance during crowded events," *ACM SIGMETRICS performance evaluation review*, vol. 41, no. 1, pp. 17–28, 2013. 29

[109] R. Ahas, A. Aasa, Y. Yuan, M. Raubal, Z. Smoreda, Y. Liu, C. Ziemlicki, M. Tiru, and M. Zook, "Everyday space–time geographies: using mobile phone-based sensor data to monitor urban activity in harbin, paris, and tallinn," *International Journal of Geographical Information Science*, vol. 29, no. 11, pp. 2017–2039, 2015. 29

[110] M. Z. Shafiq, L. Ji, A. X. Liu, J. Pang, and J. Wang, "Large-scale measurement and characterization of cellular machine-to-machine traffic," *IEEE/ACM transactions on Networking*, vol. 21, no. 6, pp. 1960–1973, 2013. 29

[111] S. Almeida, J. Queijo, and L. M. Correia, "Spatial and temporal traffic distribution models for gsm," in *Gateway to 21st Century Communications Village. VTC 1999-Fall. IEEE VTS 50th Vehicular Technology Conference (Cat. No. 99CH36324)*, vol. 1. IEEE, 1999, pp. 131–135. 29

[112] M. R. Vieira, V. Frias-Martinez, N. Oliver, and E. Frias-Martinez, "Characterizing dense urban areas from mobile phone-call data: Discovery and social dynamics," in *2010 IEEE Second International Conference on Social Computing*. IEEE, 2010, pp. 241–248. 29

[113] M. De Nadai, J. Staiano, R. Larcher, N. Sebe, D. Quercia, and B. Lepri, "The death and life of great italian cities: a mobile phone data perspective," in *Proceedings of the 25th international conference on world wide web*, 2016, pp. 413–423. 29

[114] A. Furno, R. Stanica, and M. Fiore, "A comparative evaluation of urban fabric detection techniques based on mobile traffic data," in *Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015*, 2015, pp. 689–696. 30

[115] V. Soto and E. Frías-Martínez, "Automated land use identification using cell-phone records," in *Proceedings of the 3rd ACM international workshop on MobiArch*, 2011, pp. 17–22. 30

[116] S. Grauwin, S. Sobolevsky, S. Moritz, I. Gódor, and C. Ratti, "Towards a comparative science of cities: Using mobile traffic records in new york, london, and hong kong," *Computational approaches for urban environments*, pp. 363–387, 2015. 30

[117] B. Cici, M. Gjoka, A. Markopoulou, and C. T. Butts, "On the decomposition of cell phone activity patterns and their connection with urban ecology," in *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2015, pp. 317–326. 30

[118] A. Furno, N.-E. El Faouzi, M. Fiore, and R. Stanica, "Fusing GPS probe and mobile phone data for enhanced land-use detection," in *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*. IEEE, 2017, pp. 693–698. 33

[119] I. T. Joliffe and B. Morgan, "Principal component analysis and exploratory factor analysis," *Statistical methods in medical research*, vol. 1, no. 1, pp. 69–95, 1992. 34

[120] T. Arentze, H. Timmermans, F. Hofman, and N. Kalfs, "Data needs, data collection, and data quality requirements of activity-based transport demand models," *Transportation research circular*, no. E-C008, pp. 30–p, 2000. 35

[121] F. Giannotti and D. Pedreschi, *Mobility, data mining and privacy: Geographic knowledge discovery*. Springer Science & Business Media, 2008. 35

[122] P. Bonnel, "Postal, telephone, and face-to-face surveys: How comparable are they?" in *Transport survey quality and innovation*. Emerald Group Publishing Limited, 2003, pp. 215–237. 35

[123] J. Wolf, M. Oliveira, and M. Thompson, "Impact of underreporting on mileage and travel time estimates: Results from global positioning system-enhanced household travel survey," *Transportation research record*, vol. 1854, no. 1, pp. 189–198, 2003. 35

[124] F. Wang and C. Chen, "On data processing required to derive mobility patterns from passively-generated mobile phone data," *Transportation Research Part C: Emerging Technologies*, vol. 87, pp. 58–74, 2018. 35

[125] M. Fekih, T. Bellemans, Z. Smoreda, P. Bonnel, A. Furno, and S. Galland, "A data-driven approach for origin–destination matrix construction from cellular network signalling data: a case study of lyon region (france)," *Transportation*, vol. 48, pp. 1671–1702, 2021. 36

[126] l. e. l. c. p. F. Centre d'études sur les réseaux, les transports, *L'enquête ménages, déplacements" standard Certu"*. Ministère de l'écologie, de l'énergie, du développement durable et de l . . . , 2008. 37

[127] M. Shafiei, M. Nazemi, and S. Seyedabrishami, "Estimating time-dependent origin–destination demand from traffic counts: extended gradient method," *Transportation Letters*, vol. 7, no. 4, pp. 210–218, 2015. 39

[128] A. Kuppam, R. Copperman, J. Lemp, T. Rossi, V. Livshits, L. Vallabhaneni, K. Jeon, and E. Brown, "Special events travel surveys and model development," *Transportation Letters*, vol. 5, no. 2, pp. 67–82, 2013. 39

[129] M.-H. Wang, S. D. Schrock, N. Vander Broek, and T. Mulinazzi, "Estimating dynamic origin-destination data and travel demand using cell phone network data," *International Journal of Intelligent Transportation Systems Research*, vol. 11, pp. 76–86, 2013. 39

[130] Z. Zhao, S.-L. Shaw, Y. Xu, F. Lu, J. Chen, and L. Yin, "Understanding the bias of call detail records in human mobility research," *International Journal of Geographical Information Science*, vol. 30, no. 9, pp. 1738–1762, 2016. 39

[131] M. Fekih, L. Bonnetain, A. Furno, P. Bonnel, Z. Smoreda, S. Galland, and T. Bellemans, "Potential of cellular signaling data for time-of-day estimation and spatial classification of travel demand: a large-scale comparative study with travel survey and land use data," *Transportation Letters*, vol. 14, no. 7, pp. 787–805, 2022. 39, 58

[132] M. Seppecher, L. Leclercq, A. Furno, D. Lejri, and T. V. da Rocha, "Estimation of urban zonal speed dynamics from user-activity-dependent positioning data and regional paths," *Transportation Research Part C: Emerging Technologies*, vol. 129, p. 103183, 2021. 39, 49, 51

[133] M. Janzen, M. Vanhoof, Z. Smoreda, and K. W. Axhausen, "Closer to the total? Long-distance travel of French mobile phone users," *Travel Behaviour and Society*, vol. 11, pp. 31–42, 4 2018. 42

[134] M. Saberi, H. S. Mahmassani, T. Hou, and A. Zockaie, "Estimating network fundamental diagram using three-dimensional vehicle trajectories: extending edie's definitions of traffic flow variables to networks," *Transportation Research Record*, vol. 2422, no. 1, pp. 12–20, 2014. 45, 47

[135] S. Uppoor, C. Ziemlicki, S. Secci, and Z. Smoreda, "On mobile traffic distribution over cellular backhauling network nodes," in *2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2016, pp. 726–731. 45

[136] N.-E. El Faouzi, A. Furno, L. Bonnetain, and M. Fiore, "Détermination de trajectoires à partir de données de téléphonie mobile," January 2023, FR3125197, EP4117318. 46

[137] M. Gramaglia, M. Fiore, A. Furno, and R. Stanica, "Glove: towards privacy-preserving publishing of record-level-truthful mobile phone trajectories," *ACM/IMS Transactions on Data Science (TDS)*, vol. 2, no. 3, pp. 1–36, 2021. 46, 74, 76, 77

[138] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1624–1639, 2011. 47

[139] Y. Zheng, F. Liu, and H.-P. Hsieh, "U-air: When urban air quality inference meets big data," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2013, pp. 1436–1444. 47

[140] J. Shang, Y. Zheng, W. Tong, E. Chang, and Y. Yu, "Inferring gas consumption and pollution emission of vehicles throughout a city," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 1027–1036. 47

[141] L. Ntziachristos, D. Gkatzoflias, C. Kouridis, and Z. Samaras, "Copert: a european road transport emission inventory model," in *Information Technologies in Environmental Engineering: Proceedings of the 4th International ICSC Symposium Thessaloniki, Greece, May 28-29, 2009*. Springer, 2009, pp. 491–504. 47

[142] M. Seppecher, L. Leclercq, A. Furno, T. Vieira da Rocha, J.-M. André, and J. Boutang, "Identification of aggregate urban mobility patterns of nonregular travellers from mobile phone data," *Future Transportation*, vol. 3, no. 1, pp. 254–273, 2023. 47

[143] M. Seppecher, "Mining call detail records to reconstruct global urban mobility patterns for large scale emissions calculation," Ph.D. dissertation, Université de Lyon, 2022. 47

[144] C. F. Daganzo, "Urban gridlock: Macroscopic modeling and mitigation approaches," *Transportation Research Part B: Methodological*, vol. 41, no. 1, pp. 49–62, 2007. 48, 52

[145] M. Yildirimoglu and N. Geroliminis, "Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams," *Transportation Research Part B: Methodological*, vol. 70, pp. 186–200, 2014. 48

[146] S. Batista, L. Leclercq, and N. Geroliminis, "Estimation of regional trip length distributions for the calibration of the aggregated network traffic models," *Transportation Research Part B: Methodological*, vol. 122, pp. 192–217, 2019. 48, 49

[147] G. Mariotte, L. Leclercq, S. Batista, J. Krug, and M. Paipuri, "Calibration and validation of multi-reservoir mfd models: A case study in lyon," *Transportation Research Part B: Methodological*, vol. 136, pp. 62–86, 2020. 49, 54

[148] S. Batista, L. Leclercq, and M. Menendez, "Dynamic traffic assignment for regional networks with traffic-dependent trip lengths and regional paths," *Transportation Research Part C: Emerging Technologies*, vol. 127, p. 103076, 2021. 49

[149] G. Chen, S. Hoteit, A. C. Viana, M. Fiore, and C. Sarraute, "Enriching sparse mobility information in call detail records," *Computer Communications*, vol. 122, pp. 44–58, 2018. 50

[150] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 759–770, 2008. 52

[151] L. Bonnetain, A. Furno, J. Krug, and N.-E. E. Faouzi, "Can we map-match individual cellular network signaling trajectories in urban environments? data-driven study," *Transportation Research Record*, vol. 2673, no. 7, pp. 74–88, 2019. 55, 60, 62, 69, 70

[152] W. Wu, Y. Wang, J. B. Gomes, D. T. Anh, S. Antonatos, M. Xue, P. Yang, G. E. Yap, X. Li, S. Krishnaswamy *et al.*, "Oscillation resolution for mobile phone cellular tower data to enable mobility modelling," in *2014 IEEE 15th International Conference on Mobile Data Management*, vol. 1. IEEE, 2014, pp. 321–328. 56

[153] B. C. Csáji, A. Browet, V. A. Traag, J.-C. Delvenne, E. Huens, P. Van Dooren, Z. Smoreda, and V. D. Blondel, "Exploring the mobility of mobile phone users," *Physica A: statistical mechanics and its applications*, vol. 392, no. 6, pp. 1459–1473, 2013. 56

[154] D. Bachir, G. Khodabandelou, V. Gauthier, M. El Yacoubi, and J. Puchinger, "Inferring dynamic origin-destination flows by transport mode using mobile phone data," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 254–275, 2019. 56

[155] S. Jiang, J. Ferreira, and M. C. Gonzalez, "Activity-based human mobility patterns inferred from mobile phone data: A case study of singapore," *IEEE Transactions on Big Data*, vol. 3, no. 2, pp. 208–219, 2017. 58

[156] P. Katsikouli, M. Fiore, A. Furno, and R. Stanica, "Characterizing and removing oscillations in mobile phone location data," in *2019 IEEE 20th International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*. IEEE, 2019, pp. 1–9. 58

[157] A. A. Taha and A. Hanbury, "An efficient algorithm for calculating the exact hausdorff distance," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 11, pp. 2153–2163, 2015. 60

[158] E. Akopyan, A. Furno, N.-E. El Faouzi, and E. Gaume, "Unsupervised real-time anomaly detection for multivariate mobile phone traffic series." in *ESANN*, 2021. 67

[159] L. Bonnetain, "Unlocking the potential of mobile phone data for large scale urban mobility estimation," Ph.D. dissertation, Université de Lyon, 2022. 69

[160] L. Bonnetain, A. Furno, and N.-E. E. Faouzi, "Multi-modal fine-grained map-matching of mobile phone network signaling data in urban area," in *101st Annual Meeting of the Transportation Research Board (TRB)*, 2022. [Online]. Available: https://people.licit-lyon.eu/furno/documents/trb_2022_bonnetain.pdf 69, 70, 71, 72

[161] Z. Shen, W. Du, X. Zhao, and J. Zou, "Dmm: Fast map matching for cellular data," in *Proceedings of the 26th annual international conference on mobile computing and networking*, 2020, pp. 1–14. 69, 70

[162] R. Mohamed, H. Aly, and M. Youssef, "Accurate real-time map matching for challenging environments," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 847–857, 2016. 70

[163] E. Algizawy, T. Ogawa, and A. El-Mahdy, "Real-time large-scale map matching using mobile phone data," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 11, no. 4, pp. 1–38, 2017. 70, 71

[164] F. Asgari, A. Sultan, H. Xiong, V. Gauthier, and M. A. El-Yacoubi, "Ct-mapper: Mapping sparse multimodal cellular trajectories using a multilayer transportation network," *Computer Communications*, vol. 95, pp. 69–81, 2016. 70

[165] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, 1967. 70

[166] R. Raymond, T. Morimura, T. Osogami, and N. Hirosue, "Map matching with hidden markov model on sampled road network," in *Proceedings of the 21st international conference on pattern recognition (icpr2012)*. IEEE, 2012, pp. 2242–2245. 71

[167] P. Newson and J. Krumm, "Hidden markov map matching through noise and sparseness," in *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems*, 2009, pp. 336–343. 72

[168] F. Asgari, A. Amrani, and M. Khouadjia, "Scaling time-dependent origin-destination matrix using growth factor model," in *2021 International Symposium on Computer Science and Intelligent Controls (ISCSIC)*. IEEE, 2021, pp. 51–57. 74

[169] B. Matet, E. Côme, A. Furno, L. Bonnetain, L. Oukhellou, and N.-E. El Faouzi, "A lightweight approach for origin-destination matrix anonymization." in *ESANN*, 2021. 74

[170] B. Matet, A. Furno, M. Fiore, E. Côme, and L. Oukhellou, "Adaptative generalisation over a value hierarchy for the k-anonymisation of origin–destination matrices," *Transportation Research Part C: Emerging Technologies*, vol. 154, p. 104236, 2023. 74, 77

[171] L. Sweeney, "Achieving k-anonymity privacy protection using generalization and suppression," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 571–588, 2002. 75

[172] M. Fiore, P. Katsikouli, E. Zavou, M. Cunche, F. Fessant, D. Le Hello, U. M. Aivodji, B. Olivier, T. Quertier, and R. Stanica, "Privacy in trajectory micro-data publishing: a survey," *Transactions on Data Privacy*, vol. 13, pp. 91–149, 2020. 75

[173] Y. Liang and R. Samavi, "Optimization-based k-anonymity algorithms," *Computers & Security*, vol. 93, p. 101753, 2020. 75, 76

[174] A. Machanavajjhala, D. Kifer, J. Abowd, J. Gehrke, and L. Vilhuber, "Privacy: Theory meets practice on the map," in *2008 IEEE 24th international conference on data engineering*. IEEE, 2008, pp. 277–286. 75

[175] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," in *2007 IEEE 23rd international conference on data engineering*. IEEE, 2006, pp. 106–115. 75

[176] C. Bettini, X. S. Wang, and S. Jajodia, "Protecting privacy against location-based personal identification," in *Secure Data Management: Second VLDB Workshop, SDM 2005, Trondheim, Norway, September 2-3, 2005. Proceedings 2*. Springer, 2005, pp. 185–199. 75

[177] W. Mahanan, W. A. Chaovalitwongse, and J. Natwichai, "Data privacy preservation algorithm with k-anonymity," *World Wide Web*, vol. 24, pp. 1551–1561, 2021. 75, 77

[178] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014. 76

[179] H. Kellerer, U. Pferschy, and D. Pisinger, *Knapsack Problems*. Springer, 01 2004. 77

[180] N. Maiti, P. Pathak, and B. Samanta, "An efficient algorithm for the precedence constraint knapsack problem with reference to large-scale open-pit mining pushback design," *Mining Technology*, vol. 130, no. 1, pp. 8–21, 2021. 77

[181] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, "Mondrian multidimensional k-anonymity," in *22nd International conference on data engineering (ICDE'06)*. IEEE, 2006, pp. 25–25. 77

[182] G. Cormode, M. Procopiuc, D. Srivastava, and T. T. Tran, "Differentially private publication of sparse data," *arXiv preprint arXiv:1103.0825*, 2011. 79

[183] A. Haldane, "Why banks failed the stress test," *BIS Review*, vol. 18, p. 2009, 2009. 83

[184] T. Schuermann, "Stress testing banks," *International Journal of Forecasting*, vol. 30, no. 3, pp. 717–728, 2014. 83

[185] N. Goldschlager, A. Selzer, and K. Cohn, "Treadmill stress tests as indicators of presence and severity of coronary artery disease," *Ann Intern Med*, vol. 85, no. 3, pp. 277–286, 1976. 83

[186] S. W. Baertschi, K. M. Alsante, and R. A. Reed, *Pharmaceutical stress testing: predicting drug degradation*. CRC Press, 2016. 83

[187] L. M. Zhang, L. Gao, S. Y. Zhou, R. W. Cheung, and S. Lacasse, "Stress testing framework for managing landslide risks under extreme storms," in *Workshop on World Landslide Forum*. Springer, 2017, pp. 17–32. 83

[188] J. Clarke and E. Obrien, "A multi-hazard risk assessment methodology, stress test framework and decision support tool for transport infrastructure networks," *Transportation Research Procedia*, vol. 14, pp. 1355–1363, 2016. 83

[189] W. Wisetjindawat, A. Kermanshah, S. Derrible, and M. Fujita, "Stochastic modeling of road system performance during multihazard events: Flash floods and earthquakes," *Journal of Infrastructure Systems*, vol. 23, no. 4, p. 04017031, 2017. 83

[190] B. Donovan and D. B. Work, "Using coarse gps data to quantify city-scale transportation system resilience to extreme events," *arXiv preprint arXiv:1507.06011*, 2015. 83

[191] M. Omer, A. Mostashari, and R. Nilchiani, "Assessing resilience in a regional road-based transportation network," *International Journal of Industrial and Systems Engineering*, vol. 13, no. 4, pp. 389–408, 2013. 83

[192] U. Brandes, "A faster algorithm for betweenness centrality," *Journal of mathematical sociology*, vol. 25, no. 2, pp. 163–177, 2001. 83, 124, 126

[193] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, pp. 35–41, 1977. 83, 124

[194] A. Duret, L. Leclercq, and N.-E. El Faouzi, "Data assimilation using a mesoscopic lighthill–whitham–richards model and loop detector data: Methodology and large-scale network application," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2560, pp. 26–36, 2016. 85

[195] M. J. Lighthill and G. B. Whitham, "On kinematic waves. ii. a theory of traffic flow on long crowded roads," in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 229, no. 1178. The Royal Society, 1955, pp. 317–345. 85

[196] P. I. Richards, "Shock waves on the highway," *Operations research*, vol. 4, no. 1, pp. 42–51, 1956. 85

[197] E. Henry, A. Furno, and N.-E. El Faouzi, "A graph-based approach with simulated traffic dynamics for the analysis of transportation resilience in smart cities," in *TRB 2019, 98th Annual Meeting Transportation Research Board*, 2019, p. 21p. 91

[198] E. Henry, A. Furno, and N.-E. E. Faouzi, "Approach to quantify the impact of disruptions on traffic conditions using dynamic weighted resilience metrics of transport networks," *Transportation research record*, vol. 2675, no. 4, pp. 61–78, 2021. 92, 93, 94, 109

[199] R. Jacob, K. Harikrishnan, R. Misra, and G. Ambika, "Measure for degree heterogeneity in complex networks and its application to recurrence network analysis," *Royal Society open science*, vol. 4, no. 1, p. 160757, 2017. 92

[200] B. Wang, H. Tang, C. Guo, and Z. Xiu, "Entropy optimization of scale-free networks' robustness to random failures," *Physica A: Statistical Mechanics and its Applications*, vol. 363, no. 2, pp. 591–596, 2006. 92

[201] J. Gao, B. Barzel, and A.-L. Barabási, "Universal resilience patterns in complex networks," *Nature*, vol. 530, no. 7590, pp. 307–312, 2016. 92, 94, 95, 96

[202] H. G. Ramirez, L. Leclercq, N. Chiabaut, C. Becarie, and J. Krug, "Travel time and bounded rationality in travellers' route choice behaviour: A computer route choice experiment," *Travel Behaviour and Society*, vol. 22, pp. 59–83, 2021. 93

[203] E. Henry, L. Bonnetain, A. Furno, N.-E. El Faouzi, and E. Zimeo, "Spatio-temporal correlations of betweenness centrality and traffic metrics," in *2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*. IEEE, 2019, pp. 1–10. 97

[204] H. Wang, J. M. Hernandez, and P. Van Mieghem, "Betweenness centrality in a weighted network," *Physical Review E*, vol. 77, no. 4, p. 046105, 2008. 97

[205] L. Dall'Asta, A. Barrat, M. Barthélemy, and A. Vespignani, "Vulnerability of weighted networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2006, no. 04, p. P04006, 2006. 97

[206] E. Henry, A. Furno, N.-E. El Faouzi, and D. Rey, "Locating park-and-ride facilities for resilient on-demand urban mobility," *Transportation Research Part E: Logistics and Transportation Review*, vol. 158, p. 102557, 2022. 100, 101, 103, 104, 105

[207] E. Henry, A. Furno, and N.-E. El Faouzi, "Reinforce: rapid augmentation of large-scale multi-modal transport networks for resilience enhancement," *Applied Network Science*, vol. 6, pp. 1–24, 2021. 100

[208] M. Guillot, A. Furno, E.-H. Aghezzaf, and N.-E. El Faouzi, "Transport network downsizing based on optimal sub-network," *Communications in Transportation Research*, vol. 2, p. 100079, 2022. 100

[209] D. Li, Y. Liu, Y. Song, Z. Ye, and D. Liu, "A framework for assessing resilience in urban mobility: Incorporating impact of ridesharing," *International Journal of Environmental Research and Public Health*, vol. 19, no. 17, p. 10801, 2022. 101

[210] F. Aros-Vera, V. Marianov, and J. E. Mitchell, "p-Hub approach for the optimal park-and-ride facility location problem," *European Journal of Operational Research*, vol. 226, no. 2, pp. 277–285, 4 2013. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0377221712008223 101, 103

[211] T. Andrejszki, A. Torok, and M. Csete, "Identifyingy the utility function of transport services from stated preferences," *Transport and Telecommunication*, vol. 16, no. 2, pp. 138–144, 6 2015. 103

[212] J. Krug, A. Burianne, C. Bécarie, and L. Leclercq, "Refining trip starting and ending locations when estimating travel-demand at large urban scale," *Journal of Transport Geography*, vol. 93, p. 103041, 2021. 105

[213] A. Furno, N.-E. E. Faouzi, R. Sharma, and E. Zimeo, "Graph-based ahead monitoring of vulnerabilities in large dynamic transportation networks," *PloS one*, vol. 16, no. 3, p. e0248764, 2021. 109, 124, 125, 126

[214] C. Colarusso, A. De Iasio, A. Furno, L. Goglia, M. A. Merzoug, and E. Zimeo, "Promenade: A big data platform for handling city complex networks with dynamic graphs," *Future Generation Computer Systems*, vol. 137, pp. 129–145, 2022. 110, 117

[215] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Physics reports*, vol. 424, no. 4-5, pp. 175–308, 2006. 111

[216] S. H. Strogatz, "Exploring complex networks," *nature*, vol. 410, no. 6825, pp. 268–276, 2001. 111

[217] A. Azzara, M. Petracca, and P. Pagano, "The icsi m2m middleware for iot-based intelligent transportation systems," in *2015 IEEE 18th Int. Conference on Intelligent Transportation Systems*. IEEE, 2015, pp. 155–160. 111

[218] A. Al-Dweik, R. Muresan, M. Mayhew, and M. Lieberman, "Iot-based multifunctional scalable real-time enhanced road side unit for intelligent transportation systems," in *2017 IEEE 30th Canadian conference on electrical and computer engineering (CCECE)*. IEEE, 2017, pp. 1–6. 111

[219] A. Balalaie, A. Heydarnoori, and P. Jamshidi, "Microservices architecture enables devops: Migration to a cloud-native architecture," *Ieee Software*, vol. 33, no. 3, pp. 42–52, 2016. 113

[220] L. E. Lwakatare, P. Kuvaja, and M. Oivo, "Relationship of devops to agile, lean and continuous deployment," in *International Conference on Product-Focused Software Process Improvement*. Springer, 2016, pp. 399–415. 113

[221] N. Marz and J. Warren, *Big Data: Principles and best practices of scalable real time data systems*. Manning Publications, 2013. 114

[222] J. Kreps, "Questioning the lambda architecture," https://www.oreilly.com/radar/questioning-the-lambda-architecture/, August 2014. 114

[223] S. Chouali, A. Boukerche, A. Mostefaoui, and M. A. Merzoug, "Formal verification and performance analysis of a new data exchange protocol for connected vehicles," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 15 385–15 397, 2020. 115

[224] A. Furno, N. E. Faouzi, R. Sharma, V. Cammarota, and E. Zimeo, "A graph-based framework for real-time vulnerability assessment of road networks," in *2018 IEEE International Conference on Smart Computing, SMARTCOMP 2018, Taormina, Sicily, Italy, June 18-20, 2018*, 2018, pp. 234–241. 117

[225] A. De Iasio, A. Furno, L. Goglia, and E. Zimeo, "A microservices platform for monitoring and analysis of iot traffic data in smart cities," in *2019 IEEE Int. Conference on Big Data (Big Data)*. IEEE, 2019, pp. 5223–5232. 117

[226] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, "Network analysis in the social sciences," *Science*, vol. 323, no. 5916, pp. 892–895, 2009. 124

[227] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, "Attack vulnerability of complex networks," *Physical Review E*, vol. 65, no. 5, p. 056109, 2002. 124

[228] T. Carpenter, G. Karakostas, and D. Shallcross, "Practical issues and algorithms for analyzing terrorist networks," in *Proceedings of the western simulation multiconference*, 2002. 124

[229] D. King, A. Shalaby, and P. Eng, "Performance metrics and analysis of transit network resilience in toronto," *Transportation Research Record*, pp. 16–2441, 2016. 124

[230] Y. Zhang, X. Wang, P. Zeng, and X. Chen, "Centrality characteristics of road network patterns of traffic analysis zones," *Transportation research record*, vol. 2256, no. 1, pp. 16–24, 2011. 124

[231] B. Berche, C. Von Ferber, T. Holovatch, and Y. Holovatch, "Resilience of public transport networks against attacks," *The European Physical Journal B*, vol. 71, pp. 125–137, 2009. 124

[232] A. Furno, N.-E. El Faouzi, R. Sharma, and E. Zimeo, "Two-level clustering fast betweenness centrality computation for requirement-driven approximation," in *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, 2017, pp. 1289–1294. 124, 126, 132

[233] U. Brandes and C. Pich, "Centrality estimation in large networks," *International Journal of Bifurcation and Chaos*, vol. 17, no. 07, pp. 2303–2318, 2007. 124

[234] C. Daniel, A. Furno, L. Goglia, and E. Zimeo, "Fast cluster-based computation of exact betweenness centrality in large graphs," *Journal of Big Data*, vol. 8, no. 1, pp. 1–39, 2021. 124, 125, 126, 129, 133

[235] A. Furno, N.-E. El Faouzi, R. Sharma, and E. Zimeo, "Fast approximated betweenness centrality of directed and weighted graphs," in *Complex Networks and Their Applications VII: Volume 1 Proceedings The 7th International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2018 7*. Springer, 2019, pp. 52–65. 125, 126

[236] R. W. Floyd, "Algorithm 97: shortest path," *Communications of the ACM*, vol. 5, no. 6, p. 345, 1962. 125

[237] P. Suppa and E. Zimeo, "A clustered approach for fast computation of betweenness centrality in social networks," in *2015 IEEE International Congress on Big Data*. IEEE, 2015, pp. 47–54. 126, 132

[238] C. Daniel, A. Furno, and E. Zimeo, "Cluster-based computation of exact betweenness centrality in large undirected graphs," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 603–608. 126

[239] Y. Li, W. Li, Y. Tan, F. Liu, Y. Cao, and K. Y. Lee, "Hierarchical decomposition for betweenness centrality measure of complex networks," *Scientific Reports*, vol. 7, no. 1, p. 46491, 2017. 126, 132, 135, 138, 139

[240] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509–512, 1999. 126

[241] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008. 129

[242] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *AAAI*, 2015. [Online]. Available: http://networkrepository.com 136

[243] J. Mcauley and J. Leskovec, "Learning to discover social circles in ego networks," *NIPS*, vol. 1, pp. 539–547, 01 2012. 136

[244] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution: Densification and shrinking diameters," *ACM Trans Knowledge Discov Data*, vol. 1, 04 2006. 136

[245] C. Daniel, P. Lemaire, A. Ladino, A. Furno, N.-E. E. Faouzi, and S. Hassas, "COMFORT: Cooperative multi-agent framework for large-scale routing-based traffic control," in *102nd Annual Meeting of the Transportation Research Board (TRB)*, 2023. [Online]. Available: https://people.licit-lyon.eu/furno/documents/comfort_furno_et_al_2022.pdf 140, 142

[246] J. Haddad, "Optimal perimeter control synthesis for two urban regions with aggregate boundary queue dynamics," *Transportation Research Part B: Methodological*, vol. 96, pp. 1–25, 2017. 140

[247] I. I. Sirmatel and N. Geroliminis, "Economic model predictive control of large-scale urban road networks via perimeter control and regional route guidance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1112–1121, 2017. 140

[248] L. Leclercq, A. Ladino, and C. Becarie, "Enforcing optimal routing through dynamic avoidance maps," *Transportation Research Part B: Methodological*, vol. 149, pp. 118–137, 2021. 140, 141

[249] L. Leclercq, N. Chiabaut, and B. Trinquier, "Macroscopic fundamental diagrams: A cross-comparison of estimation methods," *Transportation Research Part B: Methodological*, vol. 62, pp. 1–12, 2014. 141

[250] V. Lequay, M. Lefort, S. Mansour, and S. Hassas, "Ajustement diffus et adaptatif de la consommation électrique résidentielle par un système multi-agent auto-adaptatif," *Revue des Sciences et Technologies de l'Information-Série RIA: Revue d'Intelligence Artificielle*, vol. 31, no. 4/2017, pp. 427–447, 2017. 141

[251] J. D. Caicedo, M. C. González, and J. L. Walker, "Public transit demand prediction during highly dynamic conditions: A meta-analysis of state-of-the-art models and open-source benchmarking infrastructure," *arXiv preprint arXiv:2306.06194*, 2023. 147

[252] S. Zhu, H. Masud, C. Xiong, Z. Yang, Y. Pan, and L. Zhang, "Travel behavior reactions to transit service disruptions: study of metro safetrack projects in washington, dc," *Transportation Research Record*, vol. 2649, no. 1, pp. 79–88, 2017. 147

[253] H. Budnitz, L. Chapman, and E. Tranos, "Better by bus? insights into public transport travel behaviour during storm doris in reading, uk," *Weather*, vol. 73, no. 2, pp. 54–60, 2018. 147

[254] A. S. Benam, A. Furno, and N.-E. E. Faouzi, "Exploring the multi-modal demand dynamics during transport system disruptions," *arXiv preprint arXiv:2307.00877*, 2023. 148

[255] C. Zhang, L. Zhang, Y. Liu, and X. Yang, "Short-term prediction of bike-sharing usage considering public transport: A lstm approach," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 1564–1571. 151

[256] D. Chai, L. Wang, and Q. Yang, "Bike flow prediction with multi-graph convolutional networks," in *Proceedings of the 26th ACM SIGSPATIAL international conference on advances in geographic information systems*, 2018, pp. 397–400. 151

[257] L. Chen, D. Zhang, L. Wang, D. Yang, X. Ma, S. Li, Z. Wu, G. Pan, T.-M.-T. Nguyen, and J. Jakubowicz, "Dynamic cluster-based over-demand prediction in bike sharing systems," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 841–852. 151

[258] W. Jiang, "Bike sharing usage prediction with deep learning: a survey," *Neural Computing and Applications*, vol. 34, no. 18, pp. 15 369–15 385, 2022. 151

[259] Y. Li and B. Shuai, "Origin and destination forecasting on dockless shared bicycle in a hybrid deep-learning algorithms," *Multimedia Tools and Applications*, vol. 79, pp. 5269–5280, 2020. 151

[260] J. Ke, X. Qin, H. Yang, Z. Zheng, Z. Zhu, and J. Ye, "Predicting origin-destination ride-sourcing demand with a spatio-temporal encoder-decoder residual multi-graph convolutional network," *Transportation Research Part C: Emerging Technologies*, vol. 122, p. 102858, 2021. 151

[261] J. An, L. Fu, M. Hu, W. Chen, and J. Zhan, "A novel fuzzy-based convolutional neural network method to traffic flow prediction with uncertain traffic accident information," *Ieee Access*, vol. 7, pp. 20 708–20 722, 2019. 151

[262] K. Kim, "Investigation on the effects of weather and calendar events on bike-sharing according to the trip patterns of bike rentals of stations," *Journal of transport geography*, vol. 66, pp. 309–320, 2018. 151

[263] R. Rochas, A. Furno, and N.-E. E. Faouzi, "Empirical analysis of the forecasting accuracy of ST-ED-RMGC with bike-sharing data under atypical weather-related scenarios," in *Proceedings of the 14th International Conference on Application of Statistics and Probability in Civil Engineering (ICASP14)*, to appear. 151

[264] ——, "Contextual data integration for bike-sharing demand prediction with graph neural networks in degraded weather conditions," in *Proceedings of the 26th IEEE International Conference on Intelligent Transportation Systems (ITSC 2023)*, to appear. 151

[265] Y. Liang, G. Huang, and Z. Zhao, "Joint demand prediction for multimodal systems: A multi-task multi-relational spatiotemporal graph neural network approach," *Transportation Research Part C: Emerging Technologies*, vol. 140, p. 103731, 2022. 152

[266] J. Ji, J. Wang, Z. Jiang, J. Jiang, and H. Zhang, "Stden: Towards physics-guided neural networks for traffic flow prediction," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 4, 2022, pp. 4048–4056. 152

[267] R. Hasani, M. Lechner, A. Amini, D. Rus, and R. Grosu, "Liquid time-constant networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 9, 2021, pp. 7657–7666. 153

[268] M. Chahine, R. Hasani, P. Kao, A. Ray, R. Shubert, M. Lechner, A. Amini, and D. Rus, "Robust flight navigation out of distribution with liquid neural networks," *Science Robotics*, vol. 8, no. 77, p. eadc8892, 2023. 153

[269] S. Hörl and M. Balac, "Synthetic population and travel demand for paris and île-de-france based on open and publicly available data," *Transportation Research Part C: Emerging Technologies*, vol. 130, p. 103291, 2021. 153, 154

[270] A. Ahrens and S. Lyons, "Do rising rents lead to longer commutes? a gravity model of commuting flows in ireland," *Urban Studies*, vol. 58, no. 2, pp. 264–279, 2021. 153

[271] B. J. Vitins, A. Erath, and K. W. Axhausen, "Integration of a capacity-constrained workplace choice model: Recent developments and applications with an agent-based simulation in singapore," *Transportation Research Record*, vol. 2564, no. 1, pp. 1–13, 2016. 153

[272] A. Justen, F. J. Martínez, and C. E. Cortés, "The use of space–time constraints for the selection of discretionary activity locations," *Journal of Transport Geography*, vol. 33, pp. 146–152, 2013. 153

[273] L. Bonnetain, A. Furno, N.-E. El Faouzi, M. Fiore, R. Stanica, Z. Smoreda, and C. Ziemlicki, "TRANSIT: Fine-grained human mobility trajectory inference at scale with mobile network signaling data," *Transportation Research Part C: Emerging Technologies*, vol. 130, p. 103257, 2021. 153

[274] Cerema, "lil-1023: Enquête ménage déplacement, lyon / aire métropolitaine lyonnaise," 2015, available online: https://data.progedo.fr/studies/doi/10.13144/lil-1023. 154

[275] K. W Axhausen, A. Horni, and K. Nagel, *The multi-agent transport simulation MATSim*. Ubiquity Press, 2016. 156

[276] R. Gasper, A. Blohm, and M. Ruth, "Social and economic impacts of climate change on the urban environment," *Current Opinion in Environmental Sustainability*, vol. 3, no. 3, pp. 150–157, 2011. 156

[277] M. Salimi and S. G. Al-Ghamdi, "Climate change impacts on critical urban infrastructure and urban resiliency strategies for the middle east," *Sustainable Cities and Society*, vol. 54, p. 101948, 2020. 156

[278] A. Rehman, K. Haseeb, T. Saba, J. Lloret, and Z. Ahmed, "Towards resilient and secure cooperative behavior of intelligent transportation system using sensor technologies," *IEEE Sensors Journal*, vol. 22, no. 7, pp. 7352–7360, 2022. 156

[279] V. M. Nik, A. Perera, and D. Chen, "Towards climate resilient urban energy systems: a review," *National Science Review*, vol. 8, no. 3, p. nwaa134, 2021. 156

[280] K. T. Chui, P. Ordóñez de Pablos, C.-w. Shen, M. D. Lytras, and P. Vasant, "Towards sustainable smart city via resilient internet of things," in *Resilience in a Digital Age: Global Challenges in Organisations and Society*. Springer, 2022, pp. 117–135. 156

[281] J. Villamayor, D. Rexachs, E. Luque, and D. Lugones, "Raas: Resilience as a service," in *2018 18th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*. IEEE, 2018, pp. 356–359. 157

[282] R. Amghar, S. Jaber, S. H. M. Moghaddam, N. Bhouri, and M. Ameli, "Resilience as a service for transportation networks: Definition and basic concepts," *Transportation Research Record*, p. 03611981231170180, 2023. 157

[283] Q. Zou and S. Chen, "Enhancing resilience of interdependent traffic-electric power system," *Reliability Engineering & System Safety*, vol. 191, p. 106557, 2019. 157

[284] X. Lu, K. Hinkelman, Y. Fu, J. Wang, W. Zuo, Q. Zhang, and W. Saad, "An open source modeling framework for interdependent energy-transportation-communication infrastructure in smart and connected communities," *IEEE Access*, vol. 7, pp. 55 458–55 476, 2019. 157

[285] M. Fiore *et al.*, "Full network sensing: Architecting 6g beyond communications," *IEEE Network*, 2023. 157

[286] A. Goudarzi, F. Ghayoor, M. Waseem, S. Fahad, and I. Traore, "A survey on iot-enabled smart grids: Emerging, applications, challenges, and outlook," *Energies*, vol. 15, no. 19, p. 6984, 2022. 157

[287] F. Poinsignon, L. Chen, S. Jiang, K. Gao, H. Badia, and E. Jenelius, "Autonomous vehicle fleets for public transport: scenarios and comparisons," *Green Energy and Intelligent Transportation*, vol. 1, no. 3, p. 100019, 2022. 157

[288] J. Ferré-Bigorra, M. Casals, and M. Gangolells, "The adoption of urban digital twins," *Cities*, vol. 131, p. 103905, 2022. 157

[289] J. Mao, L. Cao, C. Gao, H. Wang, H. Fan, D. Jin, and Y. Li, "Detecting vulnerable nodes in urban infrastructure interdependent network," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 4617–4627. 157, 158

[290] A. Hussain and P. Musilek, "Resilience enhancement strategies for and through electric vehicles," *Sustainable Cities and Society*, vol. 80, p. 103788, 2022. 158

[291] R. Hadsell, D. Rao, A. A. Rusu, and R. Pascanu, "Embracing change: Continual learning in deep neural networks," *Trends in cognitive sciences*, vol. 24, no. 12, pp. 1028–1040, 2020. 158

[292] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska *et al.*, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the national academy of sciences*, vol. 114, no. 13, pp. 3521–3526, 2017. 158