



UNIVERSITY
OF TRENTO

MASTER'S DEGREE IN
DATA SCIENCE

Level of Traffic Stress and Infrastructure Network Analysis for Safe and Accessible Cycling

Supervisor

Diego GIULIANI

Co-Supervisor

Maurizio NAPOLITANO

Candidate

Leonardo VENTUROSO

ACADEMIC YEAR
2022–2023



Ai miei cari

Acknowledgments

This work would not have been possible without the support and closeness of a great group of people. I start by thanking my supervisor Prof. Diego Giuliani for his valuable contribution in the drafting phase and for the inputs received. Next, I want to thank my co-supervisor Maurizio Napolitano, who supported me throughout the analysis process, from the initial part to the final document. Working with his research group at FBK allowed me to expand and improve my technical skills in this field of research. I also want to express my total gratitude to my family, who supported me in both difficult and beautiful moments, never doubting that I could make it. Balancing full-time work with study has been one of the most demanding (and intriguing) challenges of my life, which I am sure will yield results in the near future. I want to thank my girlfriend Giulia for her constant support in this final phase of my journey. Having someone by your side who believes in you and supports you makes things much easier, even when they are not. I also want to thank the University itself for being my 'home' over these ten years during three university courses. The people I have met over these years will always remain imprinted in my memories. I want to thank my Data Science colleagues who have helped me over the years, including Gabriele Ghisleni, Lucia Hrovatin, Mara Fantinel, Aurora Maria Tumminello and Ivan De Leonardis. And so, to quote Douglas Adams, 'So long, and thanks for all the fish!'

Abstract

The LTS-BikePlan, a sophisticated tool developed for identifying safe and accessible cycling routes, plays a pivotal role in enhancing bicycle infrastructure planning. It leverages the 'Level of Traffic Stress' (LTS) concept to assess cyclist comfort and safety by integrating diverse datasets, including OpenStreetMap (OSM) data, Geographic Information System (GIS) data like the Digital Elevation Model (DEM) and data from public administrations. This integration offers crucial insights for improving the bikeability of Italian municipalities' road network, thereby supporting sustainable mobility and urbanism initiatives. The tool's effectiveness is demonstrated through case studies (Trento, Bolzano) in varying cycling conditions, highlighting its practical applicability. Informed by key research questions, the study investigates traffic stress, infrastructure, and accessibility in relation to urban accidents. It reveals that most streets and intersections experience intermediate levels of traffic stress (LTS 3), indicating barriers to cycling, particularly for less experienced riders. High traffic stress areas are often linked to dense traffic, industrial activities and challenging topography, while lower stress areas are typically found in city centers and residential zones. The study also explores bicycle network accessibility, highlighting a trend of higher bikeability in central areas compared to peripheries. The LTS-BikePlan employs a novel approach in identifying high-stress traffic gaps within low-stress zones, offering a detailed understanding of urban cycling dynamics. Through its gap analysis and predictive models, the tool identifies areas needing infrastructure improvements and assesses accident risks. The study's comprehensive analysis suggests that roads deemed low-stress do not necessarily correlate with reduced accident rates, emphasizing the importance of infrastructure quality in cycling safety. Overall, LTS-BikePlan contributes significantly to the field of urban cycling research, providing a framework for enhancing cycling infrastructure and safety. Its insights are invaluable for urban planners and policymakers in targeting interventions effectively and promoting sustainable mobility.

Keywords: Network Analysis, Level of Traffic Stress, OSM, GIS, Gap analysis, Traffic Accidents, BNA score, PUMs, Sustainable Mobility

Contents

Acknowledgements	viii
Abstract	viii
Nomenclature	x
1 Introduction	1
1.1 Background	1
1.2 Motivation	2
1.3 Objectives	3
1.4 Research Questions	3
1.5 General Methodology	4
1.6 Thesis Structure	5
2 Literature Review	7
2.1 Psychological Stress, Traffic Related Stress, Safety-Risk Perception and Accessibility	7
2.2 Bikeability: definition and measurement	9
2.3 Sustainable Mobility and Cycling	11
2.3.1 Italian and European Existing Policies	11
2.3.2 Accidents in Italy	14
2.4 Transportation Network Analysis	15
2.4.1 Graph Theory	15
2.4.2 Shortest Path Problem and Dijkstra Algorithm	16
2.4.3 Network Statistics	18
2.5 Level of Traffic Stress, Bicycle Connectivity and Gap Analysis	19
2.5.1 Introduction to Level of Traffic Stress (LTS)	19
2.5.2 LTS research and Bicycle Network Connectivity	21
2.5.3 Gap Analysis	23
2.6 Bicycle Network Analysis Score	24
2.7 Community-Based Mapping: The OpenStreetMap Phenomenon	25
2.7.1 Formal Description	25
2.7.2 OSM Data	26
2.7.3 Tagging Process and Limitations	26
2.8 Case Studies	27
2.8.1 Trento	27
2.8.2 Bolzano	28

3	Methodology	29
3.1	Data Sources	29
3.2	Tool Design, Execution and Software	33
3.2.1	Tool Design	33
3.2.2	Software	35
3.3	Data Analysis	38
3.3.1	Data Download	38
3.3.2	Level of Traffic Stress Analysis	40
3.3.3	Exploratory Spatial Data Analysis	46
3.3.4	Network Analysis	46
3.3.5	Cluster Analysis	48
3.3.6	High Stress Gap Analysis	50
3.3.7	Urban Space Quality Assessment	54
3.3.8	BNA Score	56
3.3.9	Accidents Analysis	60
3.4	Final Analysis	61
4	Results and Discussion	65
4.1	Results	65
4.1.1	Adaptation of Level of Traffic Stress Analysis	65
4.1.2	Infrastructure and Network Analysis	68
4.1.3	Qualitative Analysis	104
4.1.4	Bicycle Network Analysis (BNA)	107
4.1.5	Final Analysis	116
4.2	Discussion	126
5	Limitations and Future Research	133
5.1	Limitations and Future Research	133
	Conclusions	137
	References	153
	List of Figures	158
	List of Tables	159

Nomenclature

ADASYN	Adaptive Synthetic Sampling
BAI	Bike Attractiveness Index
BNA	Bike Network Analysis
BSI	Bike Safety Index
BSIR	Bike Safety Index Rating
BSL	Bicycle Stress Level
CRAN	Comprehensive R Archive Network
CRS	Coordinate Reference System
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DEM	Digital Elevation Model
EPSG	European Petroleum Survey Group
EU	European Union
FIAB	Federazione Italiana Ambiente e Bicicletta - Italian Federation of Environment and Bicycle
GIS	Geographic Information System
HDBSCAN	Hierarchical Density-Based Spatial Clustering of Applications with Noise
HTML	Hypertext Markup Language
IDE	Integrated Development Environment
IEI	Intersection Evaluation Index
IHS	Interaction Hazard Score
IPDC	Identify, Prioritize, Decluster, Classify
ipynb	Interactive Python Notebook
JSON	JavaScript Object Notation
KDE	Kernel Density Estimation

NOMENCLATURE

LaTeX	A high-quality typesetting system
LISA	Local Indicators of Spatial Association
LTS	Level of Traffic Stress
MRS	Margin Rates of Substitution
NaN	Not a Number
OPTICS	Ordering Points To Identify the Clustering Structure
OSM	OpenStreetMap
PSC	Perception of Safety and Comfort
PSCL	Home-Work Travel Plans
rfc	Random Forest Classifier
RSI	Roadway Segment Index
SCC	Strongly Connected Components
SMOTE	Synthetic Minority Over-sampling Technique
SMOTEENN	SMOTE and Edited Nearest Neighbors
SUMP	Sustainable Urban Mobility Plan
svm	Support Vector Machine
WCS	Web Coverage Service
WMS	Web Map Service
WMTS	Web Map Tile Service

Chapter 1

Introduction

1.1 Background

The advancement of sustainable transportation stands as a paramount global endeavor in contemporary times. This urgency is particularly pronounced in urban contexts and within the ambit of the European Union (EU), where addressing ecological concerns, augmenting public health, diversifying energy resources and elevating the quality of urban living have become pivotal. In the EU, the thrust to mitigate local air pollution, decrease traffic-related incidents, and alleviate congestion forms an integral part of its long-standing policy landscape, as evidenced by various strategic documents and initiatives [57, 97, 224]. Among the plethora of sustainable transport modalities, cycling emerges as a particularly advantageous option. It offers a spectrum of benefits ranging from diminishing emissions and fostering physical health to preserving biodiversity and improving accessibility and socialization [43, 44, 100, 116, 137, 179, 199, 249]. To effectively promote cycling as a viable and safe transportation alternative, it necessitates a thorough exploration of existing cycling infrastructures. This exploration should aim to pinpoint and address barriers that hinder the optimal experience of cycling. In Italy, the momentum towards sustainable mobility has been significantly bolstered through the implementation of pivotal strategic frameworks such as the Biciplan and Sustainable Urban Mobility Plans (SUMP; *Italian: Piano Urbanistico di Mobilità Sostenibile*). The National Cycling Plan (Biciplan) is a comprehensive strategy aimed at integrating cycling into the urban and interurban mobility matrix. It underscores the imperative for infrastructural enhancements, policy support and public awareness initiatives [84, 163]. Concurrently, SUMP represent a localized urban planning approach that concentrates on sustainable mobility. These plans include a suite of measures to refine cycling infrastructure, curtail congestion, and boost the overall efficacy of transportation systems [222, 226]. At the heart of this discourse is the notion of 'Level of Traffic Stress' (LTS) experienced by cyclists. This concept encapsulates the subjective experiences of discomfort or perceived danger that cyclists face while navigating urban road networks [86, 90, 160, 220]. LTS is a multifaceted construct, influenced by an amalgamation of factors including the design and quality of cycling infrastructure, accessibility to common destinations, the behaviors of vehicular traffic, environmental conditions and topographical characteristics of the urban landscape [90, 150]. By evaluating the level of traffic stress, urban planners and policymakers can identify critical areas for improvement, thereby fostering a safer, more inviting environment for cycling. This, in turn, has the potential to significantly influence the uptake of cycling as a sustainable and health-promoting mode of transport in urban areas [64, 231]. To our knowledge, various methods and techniques have been employed to investigate the level of traffic stress and more generally

the concept of bikeability, ranging from the analysis of more topological aspects [160] related to infrastructure to more qualitative aspects such as those inherent in the social sphere, such as accessibility to common destinations and services [150, 177, 189]. The proposed tool, named *LTS Bike Plan*, is envisioned as a comprehensive instrument for evaluating the conditions for cycling. It adopts a holistic approach, blending both qualitative and quantitative methodologies to assess the Level of Traffic Stress (LTS) and the overall bikeability of an area. This tool delves into a spectrum of factors that collectively shape the cycling experience. These factors encompass aspects such as the connectivity of the road network, the terrain's gradient, the existence and condition of dedicated cycling paths, the accessibility of essential amenities and the relationship with accidents and safety issues. By integrating these varied indicators, the *LTS Bike Plan* aims to offer a detailed and precise analysis of the current cycling infrastructure, highlighting specific regions in need of strategic improvement to enhance bikeability. A key differentiator of the *LTS Bike Plan* from existing tools is its versatility in adapting to varying topographies, complex urban layouts, and distinct regional characteristics (e.g. mountainous areas). This adaptability is achieved through the integration of Geographic Information Systems (GIS) data, Open Street Map (OSM) data, and proven specialized spatial and network analysis techniques. Other methodologies are tailored and adapted specifically to evaluate the Level of Traffic Stress and the bikeability of a city. The tool's design emphasizes replicability, ensuring its utility in a wide range of contexts. This versatility makes the *LTS Bike Plan* a valuable asset in aligning with current sustainable transportation and infrastructure development policies. It promises to be a pivotal resource in promoting cycling as a sustainable and attractive mode of transportation, not only in varied Italian contexts but also across diverse European landscapes. This aligns with broader environmental and health objectives, making it a crucial component in advancing sustainable urban mobility.

1.2 Motivation

The genesis of the *LTS Bike Plan* is deeply rooted in the pressing need to align contemporary urban mobility solutions with the stringent requirements of sustainable, safe and accessible transportation. This motivation stems from a profound understanding that the future of urban transit must not only be environmentally friendly but also inherently safe and universally accessible. The tool is designed as a response to the emerging challenges and risks associated with urban cycling, aiming to systematically address these through innovative and practical solutions. A central element of this motivation is the prevention of risks associated with cycling in urban areas. The *LTS Bike Plan* embodies a proactive approach towards identifying potential hazards that cyclists face. This approach aims at helping urbanists and policy-makers to transform urban landscapes into safer and more inviting environments for cycling, which is a crucial step towards encouraging a shift away from reliance on automobiles. Furthermore, the *LTS Bike Plan* aligns seamlessly with current policies emphasizing sustainable and safe mobility. It acts as a bridge between policy intentions and practical, on-ground improvements in cycling infrastructure and urban planning. The tool's comprehensive nature – considering factors from road network connectivity to safety (road accidents) and accessibility – ensures that it is not just a theoretical model but a practical instrument for effecting real change. The overarching goal is to create urban spaces that are not only conducive to cycling but also promote it as a desirable, safe, and accessible option. This vision aligns with broader global and regional efforts towards sustainable urban development, particularly in the context of the European Union's focus on ecological preservation and public health enhancement [222, 224, 225]. In summary, the *LTS Bike Plan* is motivated by the urgent need to develop a holistic and adaptable tool that not only

aligns with but also actively promotes current policies on sustainable, safe, and accessible urban mobility. Its focus on reducing stress and preventing risks is a testament to its forward-thinking approach, ensuring that cycling becomes a cornerstone of urban transit systems, cherished for its contributions to public health, environmental sustainability and the overall quality of urban life.

1.3 Objectives

This research delves into the multifaceted nature of transportation preference, particularly focusing on factors influencing the choice between bicycles and cars. As highlighted by Reggiani et al. [204], the decision-making process in this context is complex, integrating a spectrum of elements ranging from logistical and environmental conditions to temporal constraints and personal characteristics or inclinations. The task of gathering individual preferences to develop a universally applicable tool is notably challenging, largely due to diverse cultural contexts. For instance, the preference patterns observed among residents of Copenhagen are likely to contrast significantly with those in a city like Rome. This variability underscores the complexity inherent in collecting and interpreting data that accurately reflects a wide range of cultural perspectives. The primary objective of this study is to elucidate the relationship between traffic stress levels and risk perception, concentrating on quantifiable aspects such as infrastructural and environmental conditions (e.g. centrality measures, network gaps), along with more qualitative factors like accessibility to common destinations within a network or the presence of accidents. The aim is to leverage open-source data to understand how these relationships impact individual decision-making processes. A central focus of this research is the development of a practical, open digital tool using a Python-based infrastructure. This tool is designed to assist public administrations by providing a comprehensive overview of potential interventions to promote bicycle usage. The investigation specifically focuses on two case studies: Trento and Bolzano. This selection is driven by two primary reasons: familiarity with Trento area, enabling effective validation of the research findings, and similar geomorphological variability, which are well-suited for analysis using satellite data. Although the study concentrates on these specific locations, the nature and structure of the developed tool and the data it utilizes are designed to be applicable across various contexts within the Italian landscape.

1.4 Research Questions

The research objectives delineated earlier culminate in a series of research questions, crucial for understanding the dynamics of urban cycling networks in varied geomorphological contexts. These questions are crafted to unravel the complexities of urban cycling networks, focusing on traffic stress, accessibility and the interplay between these elements and urban accidents. The questions are as follows:

- **1. Identification and Clustering of Traffic Stress Areas:**
In urban landscapes with diverse geomorphological characteristics, which streets or areas exhibit high traffic stress, and which ones are characterized by low stress? Are there identifiable clusters of high traffic stress within the city?
- **2. Assessment of Accessibility and Proximity to Key Destinations:**
Which streets and areas in the urban landscape exhibit higher levels of accessibility and are

in close proximity to common destinations, including services and amenities? Additionally, what is the general state of accessibility across the entire area?

- **3. Analysis of Network Subgraphs, Gaps and Traffic Stress Patterns:**
Are the networks composed of subgraphs distinguished by different levels of traffic stress? If this is the case, how many high-stress gaps occur within different areas of low stress? Moreover, which of these high-stress gaps are most crucial?

- **4. Spatial Correlation Between Accidents and High-Stress Gaps:**
Does a spatial relationship exist between the locations of accidents and the high traffic stress gaps within the urban network?

- **5. Predictive Analysis of High-Risk Areas:**
Is it possible to predict areas at high and low risk for accidents by combining quantitative metrics, such as centrality measures, with qualitative assessments like accessibility scores and levels of traffic stress?

These questions aim to dissect the multifaceted nature of urban cycling networks, shedding light on how various elements interconnect and impact cyclists' decisions and safety. The answers to these queries are expected to contribute significantly to the development of more cyclist-friendly and safer urban environments.

1.5 General Methodology

To achieve a comprehensive understanding of the accessibility and safety of cycling networks in Trento and Bolzano, this study adopts a methodical approach divided into three main stages. These stages are designed to systematically address the research questions previously identified. Within each stage, there are several distinct sub-steps integral to the analysis:

Stage One: Adaptation of the LTS Scoring Process

a. Contextual Adaptation: This initial phase involves adapting the original Level of Traffic Stress (LTS) scoring process, developed by Mekuria, Furth, and Nixon [160] and Bike Ottawa [24] for U.S. and Canadian cities, to suit the unique urban and suburban morphology of Italian cities. Key adaptations address the differences in data tagging on OpenStreetMap (OSM), street types, and speed limits.

b. Technical Implementation: Utilizing Python and R scripts, the methodology reconfigures the models and functions proposed by Bike Ottawa [24], a no-profit advocacy group interested in safe and accessible transportation. The outcome is an automated architecture that processes OSM vector data and raster data from sources like Tinitaly and other Open Data Portals.

c. Results: The result is a refined network dataset for the region of interest, accompanied by visual representations of the enriched network.

Stage Two: Infrastructure and Network Analysis

a. Exploratory Geospatial and Network Analysis: This step entails an in-depth examination of infrastructure data, focusing on metrics like centrality and intersection density, as well as identifying areas of high traffic stress within the network. It also explores the relationships between these areas and varying levels of traffic stress.

2. Cluster Analysis: The methodology employs various clustering algorithms to examine the distribution of traffic stress levels across network intersections, comparing the significance of the results.

3. High Stress Gap Analysis: The objective of this step is to apply an adapted Level of Traffic Stress (LTS) methodology, as described by Vybornova et al. [237], within the Identify, Prioritize, Declustering, Classify (IPDC) framework. This approach concentrates on identifying and analyzing high-stress gaps that divide low-stress subgraphs.

Stage Three: Qualitative Analysis

a. Urban Space Quality Assessment: This step represents an initial effort to evaluate urban space elements that impact the cycling experience, such as the accessibility of Points of Interest and travel times. It involves adapting Novack, Wang, and Zipf [177]’s pedestrian-focused approach, which examines the influence of various urban features on the pleasantness of urban spaces, to a similar, bike-centric perspective.

b. BNA Score Computation: Building on this concept, the workflow for the Bike Network Analysis (BNA) score, as developed by PeopleForBikes [191], is tailored and refined for use in Italian urban environments. This scoring approach assesses accessibility to key destinations via low-stress routes. The adaptation includes integrating demographic data and information on common destinations, covering a range of sites including service, educational, and recreational locations.

c. Accident Analysis: An in-depth analysis was conducted focusing on general accidents in both cities over the past five years. This examination aimed to uncover patterns and correlations between these accidents and varying levels of traffic stress.

Final Analysis Steps

The objective of this final stage was to synthesize insights from previous analyses and outcomes to determine if there are any correlations among the results.

a. Spatial Relationship Analysis: Utilizing the Local Indicators of Spatial Autocorrelation (LISA) method [12], the study computes the spatial relationship between accidents and high-stress gaps in the network.

b. Risk Area Prediction: This analysis involves predicting risk areas using the BNA score, degree centrality values and their spatial lags. This prediction employs various algorithms, including Logistic Regression, Random Forest Classifier, and Support Vector Machine.

1.6 Thesis Structure

The thesis is divided into several distinct chapters, each delving into various aspects of the study:

- **Chapter 2: Literature Review:**

This foundational chapter offers a deep dive into various key concepts crucial to the study. It starts with exploring the definitions and implications of stress and risk perception in

the context of cycling. The chapter then transitions into discussing bikeability, including its definition and measurement criteria. A review of sustainable mobility policies in Italy and Europe sets the stage for understanding the broader regulatory and policy framework. Furthermore, the chapter encompasses a thorough exploration of transportation network analysis and statistics. Then, it mainly examines the concept of Level of Traffic Stress, providing insights into its measurement, analysis and recent related research literature. The concept of bicycle network connectivity is scrutinized from various angles and works, followed by a detailed discussion on the Bicycle Network Analysis Score. The chapter continues with an in-depth look at OpenStreetMap, covering its description, data aspects, tagging processes, and the nuances of community involvement and limitations. Finally, the chapter concludes with an introduction of the two case studies.

- **Chapter 3: Methodology:**

This chapter presents the methodologies employed in the study. It starts with the identification of data sources and elaborates on the tool design and implementation for the study. The chapter is comprehensive in its approach to data analysis, covering multiple facets like Level of Traffic Stress Analysis, Exploratory Spatial Data Analysis, Network Analysis, Cluster Analysis, Gap Analysis, Accident Analysis and the computation of the Bicycle Network Analysis Score. It also details the testing and validation process on Trento and Bolzano. The chapter concludes with a section on Final Predictive Analysis, setting the stage for applying the findings in practical scenarios.

- **Chapter 4: Results and Discussion:**

Here, the thesis presents the outcomes of the applied methodologies and case studies. It focuses on the detailed results from the case studies in Trento and Bolzano, discussing the implications and insights derived from each.

- **Chapter 5: Limitations and Future Research:**

This chapter is dedicated to critically evaluating the limitations of the current study and suggesting avenues for future research. It helps in understanding the constraints of the methodologies used and provides a roadmap for further enhancement and exploration in the field.

- **Conclusions:**

The final chapter summarizes the key findings and contributions of the thesis. It encapsulates the major insights derived from the study and reflects on the impact and relevance of the research in the broader context of sustainable mobility and bicycle network analysis.

Chapter 2

Literature Review

2.1 Psychological Stress, Traffic Related Stress, Safety-Risk Perception and Accessibility

According to [56, 63], psychological stress is an umbrella term that encompasses all the mental or emotional strains that arises when individuals perceive a situation as challenging, threatening or overwhelming. In other words, it arises when the demands of the environment exceed an individual's perceived capacity to manage them efficiently. According to the American Psychological Association [9], it is a normal reaction to everyday pressures but can become problematic when it disrupts day-to-day functioning. Stress involves changes affecting nearly every system of the body, influencing how people feel and behave. Crosswell and Lockwood [63] work highlights the fact that in stress research we can distinguish between the experience of stressful events and the individual reactions to these events, which is a subjective and personal experience. The first type, stressful events/stressors, refer to specific events that can be objectively considered disruptive or altering to normal psychological functions and can be measured more objectively (e.g losing a relative, divorce, performing bad in school) [145]. However, the stress response is subjective and includes the cognitive, emotional and biological reactions triggered by these stressors [56]. Of the first type, one of the stressors is the so called Traffic-Related Stress, which can be defined as the negative psychological and physiological responses experienced by individuals due to factors associated with vehicular road traffic [92]. It can psychologically cause anxiety and feelings of frustration, alongside a perceived loss of control. From a cognitive standpoint, it might impair decision-making abilities and impulse control. According to Bigazzi et al. [22], interest on active mobility is rising due to its alignment with goals in sustainability, public health and equity. Consequently, transportation experts are now focusing on the traveler's perception of safety and comfort because it was found that it can influences active mobility routes and modes [22, 38, 243]. Due to the reliance of perceived loss of control on self-reported data [22], there's a shift towards using physiological measurements to overcome the drawbacks of self-reporting, as suggested in studies like [22, 48]. Building on the concept of perceived loss of control, various methodologies have been developed, like the Level of Traffic Stress for evaluating cycling infrastructure [160]. According to Bigazzi et al. [22], several environmental (surface, slope, curvature, traffic control, volume, speed, proximity, wind, temperature etc.) and travel factors (motorization, clothing, purpose, constraints etc.) can occur in influencing the experience of active travelers (pedestrians, cyclists). Then, the response to them are subjective and are shaped by whether the stimuli are seen as benign, irrelevant, or stressful. One of the most significant factors affecting comfort

and safety perceptions is exposure to motor vehicle traffic [132, 243]. Increased traffic volume, elevated speeds, and heavy-duty vehicles generally lessen comfort for pedestrians and cyclists. Conversely, physical separations between travelers and vehicles, such as buffers or medians, enhance comfort. Intersection design is also crucial; shorter crossing distances and reduced conflict points contribute to greater comfort [78, 128, 132]. Also the exposure to air-pollution has a role [94]. Additionally, the physical environment (steep, hazardous surfaces etc.), including road or path surfaces, can create discomfort through physical impacts or perceived risk of a fall and unsafety [15]. Other factors such as the quality of the infrastructure, the flexibility and the complexity of the routing has been found to influence the all travel experience [243]. Of all those enunciated, one of the certainly most influential factors is the perception of safety, especially for cycling [33, 75, 107, 139, 236]. Moreover, personal factors like age, gender, and socio-economic status influence cyclists' risk perception, with novices, recreational cyclists, women, and younger riders typically being more cautious [74]. Generally, for these range of people, separated cycling infrastructure is preferred, quiet residential streets are favored over busy roads, even with dedicated bike lanes [74, 139]. Another important factor that emerges in travel/route choice is the concept of accessibility [112]. Many studies have utilized the concept of cumulative opportunities within travel time isochrones to explore the interplay between accessibility, mode choice, and land use, even with bicycles involved [83, 122, 185, 232]. An example is the work done by Saghapour, Moridpour, and Thompson [211], who created a cycling accessibility index considering the catchment area and travel distances in Australian context. Interestingly, some works have included factors of safety, comfort and convenience in evaluating bicycle accessibility, going on to see how these factors influence each other. For instance, Lowry et al. [151] integrated the concept of Bicycle Level of Service in calculating bicycle accessibility, applying a distance-based weighting function to cycleways in an Idaho community. On the other hand, [242] proposed a bikeability index, that took into consideration bicycle facility connectivity, infrastructure quality and topography in Vancouver. However, as previously discussed regarding the perception of safety, noteworthy work in terms of accessibility, safety, and stress has been conducted by Mekuria, Furth, and Nixon [160] in their Level of Traffic Stress procedure, which will be discussed in a dedicated section. Thus, to summarize the above literature, the concept of Traffic Stress in relation to safety and accessibility, serves as a comprehensive metric to assess the various challenges and discomforts a cyclist faces on a particular road segment. This definition encapsulates not only the immediate sense of unease, anxiety or peril a cyclist might experience, but also considers the broader impacts of infrastructure design, vehicular behavior, environmental elements and topographical characteristics. The implications of this definition extend beyond personal discomfort, influencing decision-making processes of cyclists. For instance, a cyclist might opt for a quicker but riskier route over a safer, albeit steeper path, or may altogether refrain from cycling in the absence of adequate facilities. A notable example is how a steep descent in a high-traffic area might escalate accident risks due to increased speeds. From a broader perspective, environment factors, especially in the context of the improvements proposed by the Sustainable Urban Mobility Plans (SUMP) [222, 225], are perceived not just as mere physical attributes affecting the enjoyment of a ride. Instead, they are dynamic elements that interact with infrastructure layout, cycle route design, traffic behavior and the surrounding environmental conditions [15, 33, 243]. This holistic understanding underpins the rationale for incorporating available factors, such as gradient, into the Level of Traffic Stress calculations as it will be done in the Methodology section. It highlights how such integration is pivotal in enhancing the overall safety, comfort and accessibility of cycling, themes that are central to the discourse on sustainable urban mobility.

2.2 Bikeability: definition and measurement

Certainly, you have encountered scenarios where you pondered, 'What is the duration of travel from my current location to destination X?' followed by considerations such as, 'Is it feasible to walk or cycle there?' Subsequently, you may have contemplated, 'Perhaps it is unadvisable due to heavy traffic, extensive distance, or the presence of uphill sections, which could result in arriving at the destination in a state of perspiration.' It is unequivocally challenging to definitively ascertain if a destination is conveniently accessible by bicycle, as this determination is influenced by a myriad of personal and environmental variables that exhibit significant variability. In prior research, the evaluation of bicycle network accessibility has predominantly concentrated on the structural attributes of the network and connectivity, utilizing indicators like completeness or count of connected components [173, 213]. However, according to [204], the evaluation of urban bicycle networks must incorporate also user-centric considerations. According to Kellstedt et al. [133], the concept of bikeability is multi-dimensional, primarily aimed at assessing how suitable a particular environment is for cycling activities. This concept integrates essential elements like safety, comfort and the ease of connecting different points, which are crucial in determining the extent to which a certain community or region is favorable for bike use. The term is used in several research works such as [117], but interestingly after more than 15 years is still not included in dictionaries. Only the work *bikeable* is and is defined as *suitable or fit for biking* and, respectively, *suitable or safe for cyclists* [26, 27]. Lowry et al. [151] defined bikeability as "an assessment of an entire bikeway network for perceived comfort and convenience and access to important destinations". Also Arellana et al. [14] identified comfort and safety as the predominant factors for bikeability indices, while Lowry et al. [151] adopt a more holistic approach, by distinguishing between the assessment of the suitability of individual routes and evaluating the overall ease and comfort of cycling across an entire city's network. This approach recognizes that the convenience and comfort of cycling in a city depend not only on the quality of specific routes but also on how these routes integrate and function as part of the broader network. As mentioned earlier, individual factors at this point play a key role that goes from preferring quickest paths to prioritizing routes with more comfort and safety or choose longer routes instead of bicycle infrastructure not perceived as safe as they should be [36, 143]. For example, one significant factor influencing an individual's decision to cycle is their perception of the risks associated with bicycling, particularly in terms of safety [200]. Several studies found that the average detour rates of cycling trips span from 8% to as high as 93%, meaning that there's a lot of diversity in cyclists' willingness regarding detours [28, 36]. A prominent study that has effectively integrated the aspects of comfort and distance in bikeability assessment is the Level of Traffic Stress (LTS) approach by Mekuria, Furth, and Nixon [160]. This study conceptualizes user preferences in terms of a pre-determined tolerance level for traffic stress. This methodology, though somewhat simplistic, is designed to align with the needs of public administrations for tools that focus more on infrastructure-level interventions rather than individual behaviors. As highlighted by Reggiani et al. [204], this approach is based on low-stress connectivity and on the assumption that a trip maintains a certain level of stress if every segment of the journey remains below that stress threshold and the detour stays within a specific limit. According to Nielsen and Skov-Petersen [175], the components of the bikeability index are characterized by their focus on infrastructure, which includes aspects such as the proximity and access to other modes of transportation, and are also defined by topological characteristics. The increasing interest in bikeability is evident from various projects, where urban planners and researchers have developed composite bikeability scores [229, 93, 143, 206]. These projects, each initiated with distinct goals, employed varied methodologies. Often, the criteria for these indices were based more on expert opinion than empirical data. As mentioned, a bikeability assessment is the so called Copenhagen

Index [229], which evaluates cities across 14 categories, each offering a score between zero and four points, plus additional points for exceptional efforts. The index assesses cities as a whole without distinguishing between different districts. The analysis includes criteria such as bicycle culture, facilities, infrastructure, bike-sharing programs, gender split in cycling, modal splits, safety perception, political support for bikeability, social acceptance of cyclists, urban planning, traffic calming measures and the use of freight bicycles in logistics. A third example is the ADFC Fahrradklimatest conducted biennially by the Allgemeiner Deutscher Fahrradclub [4], which is a large-scale survey assessing the bicycle-friendliness of 539 cities. Qualitative data are prioritized using an online questionnaire covering aspects like cycling and traffic climate, safety, family friendliness, comfort, infrastructure and the status of bicycle traffic that are rated by respondents. A fourth example is the Bikeability Index Dresden [93] that maps bikeable areas in a city by using a score for each grid cell of the map. The value is obtained by summing values assigned to various criteria, representing the cell's bikeability. Key components of this index include bicycle infrastructure, separate bike paths, green spaces, topography, land use, bike facilities and traffic load. According to this method, factors such as steep slopes can deter cycling due to increased effort. However, each criterion is equally weighted in this case study, which might not fully capture the relative importance of each aspect. A fifth example is the Bike Score [25], which is a prominent index that offers a quantitative assessment of bikeability. It rates locations on a scale from 0 to 100 based on four key factors: the presence of bike lanes, steepness, accessibility to destinations, road connectivity and the share of bike commuting. This tool has been utilized in research to examine its correlation with economic and demographic factors and has been shown to influence cycling-to-work behaviors in various American and Canadian cities. A sixth index was proposed by Grigore et al. [101], where cyclists seek a balance between the shortest route and the highest quality of streets and intersections when choosing their path. Ease and suitability of reaching key destinations by bike is valued. This approach blends both distance and qualitative aspects of the cycling experience to assess bikeability. Kamel, Sayed, and Bigazzi [131] developed the Bike Composite Index, an amalgamation of two distinct sub-indices: the Bike Attractiveness Index (BAI) and the Bike Safety Index (BSI). This composite index integrates elements of both attractiveness and safety. Lastly, Schmid-Querg, Keler, and Grigoropoulos [212] developed a bikeability index for a specific Munich city district, focusing on criteria like bike path presence, speed limits, bicycle parking and intersection quality. Due to already mentioned OpenStreetMap data problems, the study incorporated local knowledge and field observations. This approach prioritized accuracy in a specific area over broader generalizability. Indeed McNeil [158] affirms the significance of various environmental aspects in bikeability. These include the condition of the riding surface, the availability of designated bike lanes, the presence of barriers separating cyclists from other traffic and the terrain's incline. Overall, according to an important list of scientific works, the concept of bikeability tends to be defined with a focus on specific attributes of cycling networks, such as accessibility or comfort, rather than a more comprehensive approach [80, 138, 158, 195, 211, 238, 242]. Bikeability, according to Castañon and Ribeiro [47], can guide the hierarchy of cycle network implementation, influencing evaluation, planning, and infrastructure projects. It's crucial to measure bikeability at various stages of the planning process to align infrastructure supply with cycling demand and promote bike use, ensuring alignment with local contexts.

2.3 Sustainable Mobility and Cycling

2.3.1 Italian and European Existing Policies

Mobilità Sostenibile [165] adopts the definition of 'Sustainable Mobility' from the 2006 European sustainable development strategy, focusing on aligning transportation systems to societal needs across economic, social, and environmental dimensions, while minimizing negative impacts. This concept is central to environmental policies globally and in Italy, where road transport's contribution to greenhouse gases and pollution is notable. The General Directorate for Climate and Energy in Italy spearheads sustainable mobility, aiming to mitigate transport emissions, consistent with national and EU objectives, through strategic collaborations and addressing urban mobility challenges like congestion, pollution and accidents. Several measures has been adopted in Italy to promote sustainable mobility. Recently, for example, Article 229, section 4, of the Italian Decree-Law No. 34 of May 19, 2020, mandates that enterprises and public administrations with over 100 employees in regional capitals, metropolitan cities, provincial capitals or municipalities with populations exceeding 50000, must implement Home-Work Travel Plans (PSCL) to reduce the use of private vehicles and appoint a Mobility Manager. This regulation aims to permanently lessen environmental impacts from urban and metropolitan traffic by promoting organizational and management interventions to reduce reliance on private motorized transport for daily commutes and alleviate traffic congestion [126, 165]. More interestingly, in order to improve the urban mobility system, which is an essential goal both to contribute to the fight against climate change and to counteract local scale air pollution phenomena, several European policies as been implemented [222, 225]. The European Union asserts that to meet ambitious 2050 greenhouse gas reduction targets a significant shift in both personal and goods mobility habits is essential. This entails a strategic realignment of planning towards environmental objectives, along with the promotion of innovative sustainable mobility initiatives. As mentioned above, Italy is one of the European countries with the highest motorization rate and the excessive use of private cars [165]. The Sustainable Urban Mobility Plan (SUMP) emerges as a pivotal and strategic tool for Local Authorities to address transportation and mobility challenges. It is designed to facilitate and manage the necessary transitions towards more sustainable urban transport systems [222]. Adhering to the SUMP framework [222] involves integrating mobility planning into urban design with a sustainability focus, prioritizing pedestrian and cycling mobility. It also emphasizes participatory urban mobility planning, involving various stakeholders and creating platforms for dialogue. Additionally, it advocates for spreading sustainable mobility awareness across all age groups and leverages technology in a smart city context, using it as a mean to achieve goals rather than as an end in itself. This plan encompasses all transportation modes, covering public and private transport, passenger and freight, as well as non-motorized means, traffic, and parking, thereby building on and expanding the scope of existing plans. The benefits of the Plan goes from improving the quality of life to improving accessibility and the fluidity of mobility. In the Italian national context, the Ministry of Infrastructure and Transport published a decree in August 2017 identifying the guidelines for sustainable urban mobility plans, which were updated in August 2019. The decree makes the adoption of the SUMP mandatory for cities with over 100000 inhabitants, with exceptions for those in a metropolitan city that has defined its own PUMS. Complete guidelines for developing and implementing a sustainable urban mobility plan can be found in the original work [226]. Specifically, regarding bicycle mobility, there's a project inside the SUMP known as Biciplan [163] which is a strategic plan promoting urban cycling in Italy and identifying key cycling routes in urban areas. Article 6 of the Law 11 January 2018 requires metropolitan cities and non-metropolitan municipalities to adopt the Biciplan as part of their Sustainable Urban Mobility Plan (SUMP), governed by D.M. 397/2017

and amended by D.M. 396/2019. According to D.M. 397/2017, the formulation of the Biciplan is mandatory for metropolitan cities and municipalities or associations of municipalities with a population exceeding 100000. For municipalities with a population less than 100000, they can draft this plan, but it is not compulsory. The Biciplan aims to enhance urban cycling in Italy by focusing on several key areas. These include increasing bicycle usage to balance mobility, reducing accidents among vulnerable groups like pedestrians and cyclists, cutting mobility costs, and fostering cycling as a mode of transport. The plan also seeks to develop cycle tourism and enhance the cycling network's appeal, continuity and recognizability [13, 21, 55, 58]. Regarding cycling routes, the plan categorizes them based on their function and location. The primary network connects various city parts and key sites, while the secondary network links neighborhoods to suburbs and other attraction points. Green cycling routes are designed to link parks and rural areas with the primary and secondary networks. The Biciplan not only establishes various cycling routes but also enhances the cycling experience by providing amenities like signage, bicycle parking, rental services, bike-sharing facilities, e-bike charging stations, information points, and repair shops [84]. For example, the Biciplan for Bologna (Italy) includes specific technical recommendations to enhance bicycle infrastructure and promote safe cycling. Key suggestions involve utilizing sidewalks for cycling in both urban and suburban areas to improve safety, especially where formal bike lanes are not feasible. Coloring pavement of bike lanes is proposed to increase visibility at crossings and complex areas. The plan also addresses mixed vehicular and bicycle routes, emphasizing the importance of clear markings for cyclist safety and awareness. Furthermore, it considers various aspects of urban cycling infrastructure, like bicycle lanes alongside parking areas, two-way cycling on streets, bicycle crossings, and traffic circles, all aimed at creating a more cyclist-friendly environment [21]. Another example is the Biciplan for Firenze (Italy), which is comprehensive, focusing particularly on reducing urban cycling incidents. It incorporates regulatory frameworks and technical details for implementation. Key features include advanced bicycle stop lines for improved cyclist visibility at intersections, especially when motor vehicles are turning right. It also introduces advanced bicycle houses to assist left-turn maneuvers for cyclists, which can be challenging from a right-side bike lane. The plan further suggests dedicated bike lanes or reserved lanes on roadways to enhance cyclist safety and mobility [13]. Then, regarding Trento, the document emphasizes its alignment with European and Italian laws [58]. It pays close attention to various national decrees, notably specifying guidelines like a minimum bicycle lane width of 1.5 meters. Additionally, it mandates that barriers separating bike lanes from motor vehicle lanes must be at least 0.50 meters wide. The document bears similarities to Firenze's approach. A notable point is the emphasis on the safety benefits of 30 km/h speed limits in urban areas similar to Bologna, which are seen as effective in reducing fatal and serious accidents. The analysis identifies several key problems:

- The current bike paths were designed prioritizing car traffic, leading to bike facilities being allocated wherever space was available, rather than where needed.
- The cycling network is disjointed, lacking consistency and safety.
- There is a deficient link between the city and nearby hills.
- Many sidewalks are dual-use for cyclists and pedestrians, which lessens their appeal and forces cyclists onto roads, increasing their risk.
- Bike and pedestrian paths are often narrower than regulations stipulate and are cluttered with obstacles like urban furniture, trees and manholes.
- There is a lack of areas with restricted traffic and 30 km/h zones. In some existing zones, only signs are used without actual traffic calming measures.

- There is no clear signage strategy for the cycling network, making it difficult to navigate.

The '*Guidelines for the Design of Bicycle Paths in Trento*' section stresses the need to fill gaps in the cycling network, aiming for a seamless, uniform and safe infrastructure. It suggests revising existing paths, particularly those adjacent to major roads, to reduce conflicts between pedestrians and cyclists, possibly by repurposing spaces from motor traffic lanes or parking areas. The document recommends several strategies for improvement. In urban settings, it's advised to build bike paths or one-way bike lanes on the right side of the road, coupled with 30 km/h speed limits and traffic calming measures. Current shared paths should remain for slower users like the elderly and children, while establishing separate or one-way bike lanes for quicker cyclists. When road width is insufficient for one-way bike paths, the document proposes various measures to reclaim space, such as eliminating superfluous road areas, reconfiguring parking and changing traffic flow. Continuous upkeep of shared paths is emphasized, along with the removal of obstructions and clear signage. Cycle paths on sidewalks are only recommended for continuity with existing paths in areas with low pedestrian traffic, adhering to minimum width standards. The creation of new mixed-use paths in urban areas is discouraged, favoring traffic calming and one-way bike lanes on roads. In rural areas, the preference is for bi-directional paths in protected spaces and shared paths with distinct cycling and pedestrian areas. The use of unpainted pavement for bike paths is suggested, reserving color for high-risk areas like intersections, evaluated individually. Another important document is the guide provided by *Guida per Progettare la Ciclabilità Sicura* [103] that focus on the construction of efficient and secure bicycle lanes. A critical aspect highlighted is the immediate consideration of the regulatory framework before developing solutions. This approach provides a comprehensive understanding of the current legal context, offering insights into the interpretation of 'Decreto Legge' (pages 9-12). The guide includes vital details about lane specifications such as dimensions depending on traffic or street type (e.g., 1.5 meters width for standard bike lanes) and road curb requirements, establishing ideal standards for comparative analysis. It also addresses the design of lanes in extra-urban areas. The guide proposes specific design criteria for different road contexts:

1. On roads with paved sidewalks of at least 50 cm and standard module lanes, a narrow bicycle lane is proposed, varying from 100 to 50 cm based on the motor vehicle lane width (375 cm or 325 cm). The design aims to maintain a 275 cm gap between the bike lane and the median.
2. For roads lacking sidewalks or having unpaved sidewalks, the bike lane can be a minimum of 80 cm wide from the carriageway's edge if it has flush edges, or 120 cm for vertical edges. The motor vehicle space should not be less than 220 cm, requiring a total road width of at least 600 cm.
3. On roads narrower than 600 cm, only ground markings are advised, supplemented by additional signage and speed control. On low-traffic roads primarily for cycling, the use of a central bidirectional lane of 2.5/3 m width, bordered by two lateral bike lanes marked by dashed lines, is recommended after reducing speed limits.

In concluding my review, I considered the government's guidelines for formulating a Biciplan for urban areas [164]. The Biciplan is envisioned as a long-term initiative and this approach necessitates the development of strategies that address both immediate and future needs. Emphasis is also placed on structuring the bicycle network hierarchically within the city, aiding in the delineation of different implementation phases at the local level. The guidelines underscore the need for comprehensive analysis within the Biciplan, encompassing various elements such as the evolution of demand over time, potential mobility demand under specific conditions

and socio-demographic profiles of users. It also involves examining the spatial distribution of this demand, including trip origins, destinations, and major attraction points, along with the purpose of these trips and the modes of transportation employed. A significant portion of the document advocates for the use of Geographic Information System (GIS) tools to construct a detailed representation of the cycling network. This includes associating each route or segment with its technical and functional characteristics to gain an in-depth understanding of the mobility services offered. Working within the PUMS framework provides a structured method for monitoring and evaluating the current state of cycle paths. The guidelines elaborate on assessing factors such as accessibility to crucial urban nodes and road safety, with a particular focus on accident analysis. Despite the mention of utilizing ISTAT data for micro-mobility analysis, the document notes the lack of specific useful data in this area.

2.3.2 Accidents in Italy

As mentioned inside sustainable mobility policies and different Biciplans [13, 21, 163, 226] road accidents are one of the things to adjust to enhance sustainable mobility. Indeed, accidents bring about social and economical costs [233]. According to *La Statistica ACI-Istat-Sintesi Dello Studio* [140], the year 2022 is marked by a significant rebound in mobility and, as a consequence, also in road accidents, following the years in which the pandemic was at its most severe. Compared to 2021, there is an overall increase in accidents and injuries. In 2022, Italy recorded 3159 road accident fatalities, an increase of 9.9% compared to the previous year. There were 223475 injuries (+9.2%) and 165889 road accidents (+9.2%). Although these figures show an increase from 2021, they still represent a decrease when compared to 2019, with accidents and injuries down by 3.7% and 7.4% respectively. Fatalities increased in 2022 for all road users compared to 2021, except for cyclists and occupants of trucks. There were 1375 fatalities among car occupants (+15.4%), 781 among motorcyclists (+12.4%), 70 among moped riders (+4.5%), and 485 among pedestrians (+3.2%). For truck occupants, there were 166 deaths, a decrease of 1.8%. Meanwhile, fatalities for bicycles and electric bicycles totaled 205, down from 220 in 2021 (-6.8%). However, injuries among electric scooter users increased. Specifically for electric scooters (counted since 2020), road accidents involving them rose from 2101 in 2021 to 2929 in 2022. Among incorrect driving behaviors, distraction, failure to yield right of way and excessive speed remain the most frequent. This is highlighted also by Cardamone, Eboli, and Mazzulla [45], which in a research found that Italian drivers' attitudes and behaviors impact their perception of accident risk. Specifically, the adherence to or violation of driving regulations (such as breaching safety distances or speed limits) and proper or improper driving behavior (like driver distraction) significantly influence drivers' perceptions of the likelihood of accidents. Moreover, according to Prati et al. [197], compared to cyclists of other European countries, Italian cyclists tend to report lowest ratings on infrastructure quality and provisions, so infrastructure quality cannot be underestimated in accidents discussion. Gender differences are also found in traffic danger perception [7]. Indeed, Prati et al. [197] showed that bicycle infrastructures that minimize the interaction with motorized vehicles could be beneficial for the safety of women cyclists. These three categories together account for 38.1% of cases, a figure that has remained stable over time. Interestingly, regarding active mobility, bicycle use and travel via micromobility devices have increased their share of overall transportation, rising from 3.3% in 2019 to 4.7% in the first half of 2022. Regarding the sale of traditional bicycles, after two years of booming sales, there's a slight decline, yet a total of 1.772 million bikes were sold.

2.4 Transportation Network Analysis

Transporting can be defined as the act of navigating geographic space between two points, encompassing an infinite array of intermediate locations [2]. Delving deeper into this spatial aspect, transport systems can be viewed through the lens of graph theory [23]. In the context of the increasing complexity of modern transport networks, it provides a robust framework for understanding and navigating the intricate web of transportation networks that are vital to the functioning of contemporary societies.

2.4.1 Graph Theory

In graph theory, networks are conceptualized as collections of nodes (representing locations, intersections, or stops) and edges (representing routes, paths, or connections) (see Fig. 2.1). This approach allows for a quantitative and qualitative assessment of transportation systems, focusing on properties like connectivity, accessibility, network density, and complexity. By applying graph theory, transport analysts can model and analyze these networks, optimize routes, identify critical nodes, and assess the impact of network changes or disruptions [31].

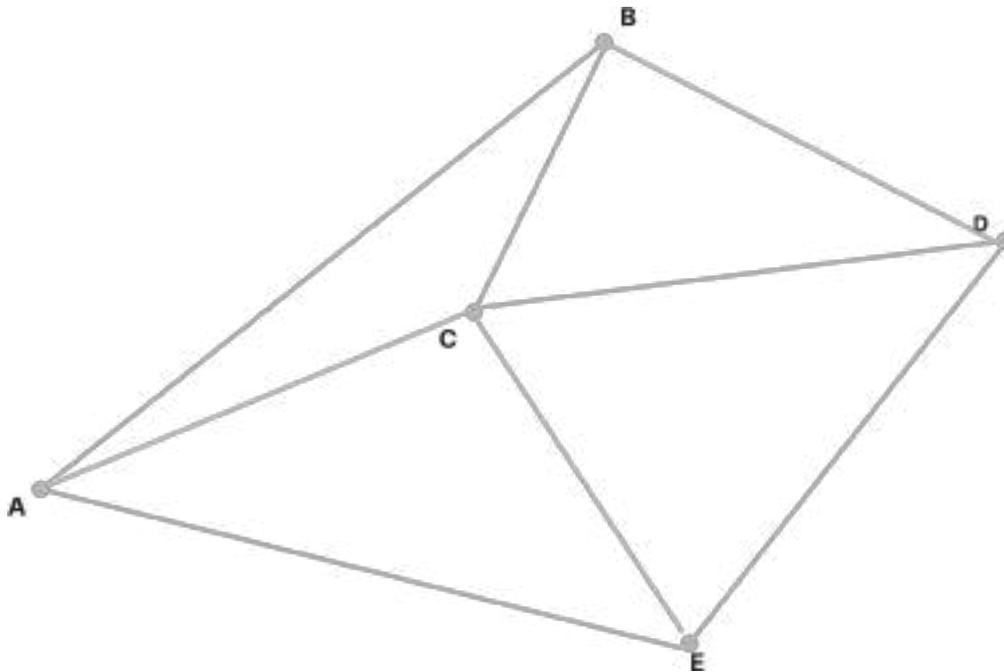


Figure 2.1: Graph representation. Each node is associated with a letter.

Additionally, the application of graph theory in transportation extends to various facets such as traffic flow optimization [66] and public transit system design [72]. For instance, in managing urban traffic, graph theory helps in optimizing signal timings [8] and traffic flow patterns to reduce congestion [3]. Concerning network resilience, graph theory assists in identifying vulnerability points and strategizing on ways to fortify the network against disruptions, whether due to natural disasters, accidents or other unplanned events (for example [217]). According to Phillips, Schwanghart, and Heckmann [192], there are several relevant terms in graph theory. As mentioned above, the two fundamental elements are called nodes and edges. A node/vertex,

is represented visually by a dot or a filled circle. Connections between nodes in a graph are established through edges/links. These can be undirected or directed. With a graph called G , a subgraph is essentially a smaller graph derived from this graph. It consists of a selection of nodes and edges, where each node and edge in the subgraph is also found in the larger graph G . The degree of a node refers to the total count of edges connected to it. In the context of a directed graph, this concept splits into in-degree and out-degree. The in-degree represents the count of edges arriving at a node, while the out-degree signifies the number of edges departing from it. Other useful terms are path and cycle. The first one is a route from one point to another where each node in between is visited once. A cycle is a path that starts and ends at the same point. Finally, a network in graph theory is a type of graph where specific attributes are linked to its nodes and edges (for example: a weighted directed graph).

2.4.2 Shortest Path Problem and Dijkstra Algorithm

The shortest path problem is a fundamental concept in Graph Theory and computer science, involving the determination of the shortest or least costly path between two points in a network, that can be represented as a graph [6, 214]. The concept of shortest is connected to various metrics that vary depending on the context (distance, time, cost etc.). The goal is to find the path that minimizes this measure from a starting node (origin) to an ending node (destination). The shortest path problem has widespread applications such as finding the shortest route for traveling between two point of a city or in telecommunications such as establishing efficient data routing (creating a route from one node to another) and urban planning [214]. To be more specific, the *Single-Source Shortest Path* determines the shortest paths from a single source node to all other nodes in the graph. A classic example of this Dijkstra's algorithm [73], which efficiently finds the shortest paths from one node to every other node in a graph with non-negative edge weights. Another one is the so called *Single-Destination Shortest Path*, which is very similar to the single-source, but focuses on finding the shortest paths to a single destination node from all other nodes. Then we have the *Single-Pair Shortest Path*, that finds the shortest path between a specified pair of nodes. This is a common requirement in routing and navigation systems. Finally, the *All-Pairs Shortest Path* determines the shortest paths between all pairs of nodes in the graph. Algorithms such as Floyd-Warshall are designed for this purpose [87]. The complexity of the problem and the best algorithm to solve it can vary depending on the characteristics of the graph (e.g., whether it's directed or undirected, weighted or unweighted, sparse or dense) and the specific requirements of the task (e.g., finding the shortest path for one pair of nodes vs. all pairs). To provide a more theoretical definition, Dijkstra's algorithm [73] is a famous and efficient algorithm used to find the shortest path from a single source node to all other nodes in a weighted graph. The graph can be either directed or undirected, but the weights must be non-negative. It starts with a source node with distance set as 0 and to all nodes as infinity. Then, a set (or priority queue) is created to track nodes whose shortest distance from the source is not yet finalized. The algorithm proceeds iteratively, selecting the node with the smallest known distance from the source node from the set. After the selection, the tentative distances to its neighboring are calculated and for each node this tentative is the sum of the distance from the source to the current node and the weight of the edge connecting the current node to the neighbor. If this tentative distance is less than the previously recorded one, the algorithm updates the shortest distance for that neighbor and the node selected is marked as 'visited' and then removed from the list. The process continues until the last node is reached. The approach is called greedy because you choose the next node with the smallest known distance. The step is repeated until the set is empty. The time complexity of Dijkstra's algorithm depends on the implementation. Using a simple array, the time complexity is $O(V^2)$,

where V is the number of vertices. With a priority queue, it can be reduced to $O(V + E \log V)$, where E is the number of edges.

I will now provide an example in order to show how the the Dijkstra's algorithm works (see Fig. 2.2).

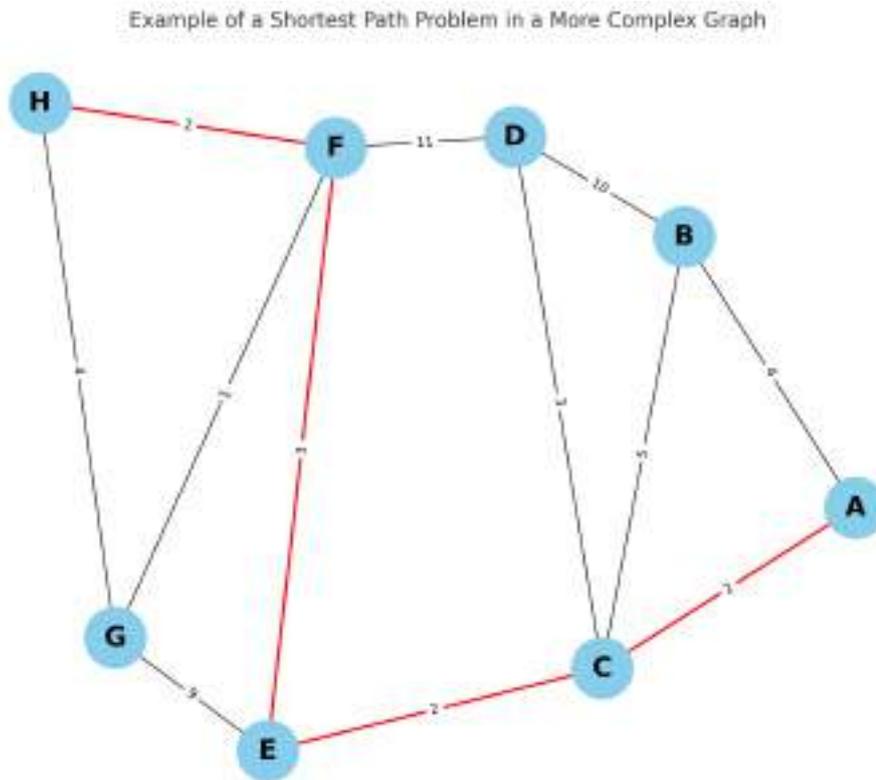


Figure 2.2: Example of a graph in order to show how Dijkstra's algorithm works

The figure can represent a road network connecting different towns or different areas of a city. Every two nodes are connected by a path, so this graph can be considered as a connected graph. Each edge has a number marked next to it, which indicates the 'cost' of traversing this edge. The cost can refer to the road length, the time etc. Therefore, the graph is a connected weighted graph. In this sense, the problem translates into finding a path from A to G with minimum total weight. By starting from the point A, we move to C (weight = 2) because is the one with the lowest cost. Then from C we move to E because the weight is again 2, next we proceed to F because the weight is 3 and finally to H which has a weight of 2. Then by summing the total costs the weight of the path A-C-E-F-H is 9. This is the shortest cumulative weight to connect the two nodes. The path is represented in red (see Fig. 2.2). In solving this shortest path problem, the algorithm (like Dijkstra's) evaluates multiple paths, calculating the cumulative weight for each and selecting the one with the lowest total weight. This process involves considering the individual weights of each edge along a potential path and making decisions at each node about which edge to traverse next to minimize the overall distance or cost.

2.4.3 Network Statistics

Network statistics in the context of street/urban networks involve a range of measures and concepts that help to analyze, classify, compare and understand the structure and characteristics of street layouts [35]. In this section, we exclusively discuss the analytical tools that are fundamental to our research.

- Degree distribution: this is a measure used to understand the connectivity of intersections (or nodes) in a street network. In simple terms, the degree of a node in a street network indicates the number of streets (or edges) connected to that node. Degree distribution is a summary of how these connections are distributed across all nodes in the network. It helps in understanding whether a street network is uniformly connected or if there are nodes with significantly higher connections than others. It serves as a validation step [237].
- Degree centrality: Degree centrality measures the importance of a node based on how many connections (or edges) it has. In simple terms, it's a count of how many direct, immediate connections a node has to other nodes in the network. For street networks, an intersection with a high degree centrality would be one connected to many streets. The formula for degree centrality for a node v is given by $C_D(v) = \frac{deg(v)}{N-1}$, where $deg(v)$ is the degree of v , and N is the number of nodes in the network.
- Closeness centrality: This measures how close a node is to all other nodes in the network, based on the shortest paths. It can help identify nodes that are most accessible within the network. The idea is that nodes with lower total distances to all other nodes are more central. The formula for closeness centrality of a node v is $C_C(v) = \frac{N-1}{\sum_{u=1}^{N-1} d(v,u)}$, where $d(v,u)$ is the shortest path distance from v to u .
- Betweenness centrality: this measures how often a node or edge acts as a bridge along the shortest path between two other nodes. It is useful for identifying streets or intersections that serve as critical connectors within the network. For street networks, an intersection with high betweenness centrality would be one that frequently appears on the shortest routes between various pairs of locations. The formula is $C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$, where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v .
- Edge Betweenness centrality: this is similar to betweenness centrality but applies to edges instead of nodes. It measures the frequency with which a given edge lies on the shortest paths between pairs of nodes. In the context of street networks, a road segment with high edge betweenness centrality would be one that is frequently used in the shortest paths traversing the network. The calculation is analogous to betweenness centrality but for edges: $C_B(e) = \sum_{s \neq t} \frac{\sigma_{st}(e)}{\sigma_{st}}$, where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(e)$ is the number of those paths that include edge e .
- Connectivity: in the context of street networks [149], connectivity refers to how well parts of the network are linked to one another. It is a measure of how easily one can navigate from one point to another within the network. In mathematical terms is the ratio of the number of existing links to the number of existing nodes. High connectivity in a street network means that there are multiple paths between locations, facilitating easier movement. It's often analyzed in terms of directness of routes, redundancy of paths and the density of the network. Pertaining to the aforementioned topic, the notion of edge connectivity emerges, characterized as the minimum quantity of links that, when removed,

results in an augmentation in the count of disconnected segments within the network. This concept serves as an indicator of the robustness of the connections among nodes in a network, gauged by the requisite number of link eliminations needed to sever the connectivity between two nodes, which can be either randomly selected or specifically identified [35, 82].

2.5 Level of Traffic Stress, Bicycle Connectivity and Gap Analysis

This section, drawing on the foundational work of Furth, Mekuria, and Nixon [90] and Mekuria, Furth, and Nixon [160], applies principles from graph theory to the analysis and categorization of bicycle infrastructure elements within urban street networks. This analysis focuses on the relationship between cyclists' tolerance for traffic-induced stress and the resulting influence on route selection. The section culminates in a comprehensive review of current advancements in network analysis as applied to bicycle networks, with a particular focus on Low Traffic Stress (LTS) research. It specifically delineates the characteristics of low-stress and high-stress bicycle networks.

2.5.1 Introduction to Level of Traffic Stress (LTS)

Literature on street network planning generally divide all the methods in two, specifically the *Demand-Driven Planning Methods* and the *Supply-Driven Planning Methods*. The first group relies on current/potential travel behavior, primarily by examining and quantifying the movement of cyclists between different areas [142]. These techniques help pinpoint enhancements that will serve the largest number of cyclists by emphasizing demand over roadway conditions. However, they fail to determine the specific improvements needed at particular locations [121]. Furthermore, demand is not static and likely depends on the conditions encountered by cyclists [142]. On the other hand, the second group concentrate on evaluating the existing road network conditions for cyclists and typically prioritize upgrades in the worst areas, leading to the establishment of protected infrastructure along all major arteries and collector roads [142]. Among supply-driven models, the *Level of Traffic Stress (LTS)* indicator has gained prominence due to its straightforward calculation and easily accessible data [160]. Indeed, several studies suggest that cyclists' route choices are influenced by the stress levels they experience on the road network [90, 210]. While supply-driven planning approaches vary in their indicator calculation methods, they all recognize a range of factors that can impact cyclists. Based on the variables considered, supply-driven approaches can also be grouped into three primary categories, focusing on:

- road and traffic conditions
- perceived factors
- physiological factors

Methodologies focusing on road and traffic conditions categorize road segments (and occasionally intersections) into distinct groups based on chosen variables with similar attributes. The *Davis Bicycle Safety Index* [69] was among the earliest systematic efforts to assess roads' operational conditions for cyclists. It computes the *Roadway Segment Index (RSI)* and *Intersection Evaluation Index (IEI)* to create the *Bicycle Safety Index Rating (BSIR)*, founded on relatively simple-to-measure road attributes. Landis [142] formulated the *Interaction Hazard Score (IHS)*, which delineates six levels of service categories employing existing road and traffic variables.

Finally, *Level of Traffic Stress (LTS)*, introduced by Mekuria, Furth, and Nixon [160], defines four traffic stress categories based on road characteristics' threshold criteria, with the lowest performing criterion determining classification. Specifically, the initial introduction of the idea of stress levels for cyclists dates back to 1978, credited to the efforts of the Geelong Bikeplan Team in Australia [220]. This concept aimed to evaluate the appropriateness of roadways from the perspective of cyclists. The team first identified and prioritized the three most significant factors impacting cyclists' stress levels: the width of the curb lane, the speed of motor vehicles, and the volume of traffic, then they developed a ranking system for various combinations of these factors, where a score of 1 represented a very low stress level and a score of 5 indicated a very high stress level. Then, Sorton and Walsh [220] put the emphasis again on 'stress' twenty years later by developing the *Bicycle Stress Level (BSL)* that classifies road segments into one of five stress levels, from very low (level 1) to high (level 5), namely not suitable for bicycle use, depending on the user. Then, the classification evolved into one more based on the attitudes of the riders towards bicycle infrastructures and stress. According to Geller [95] there are four types of cyclists in Portland, Oregon (US), namely: strong and fearless (<1%), enthused and confident (7%), interested but concerned (60%), and no way, no how (33%). The same classification has been adopted also by Dill and McNeil [75]. As mentioned above, Mekuria, Furth, and Nixon [160], based on Geller's work, developed the concept of Levels of Traffic Stress (LTS). Indeed, initially the population was classified according to their tolerance levels for traffic stress. 'Interested but concerned' category was expanded, creating two distinct subsets: one specifically for children and another for adults. For simplicity, the segment identified as 'No way, no how' was excluded from their analysis. Consequently, their assessment focused on four distinct cyclist categories. The next step was to systematically classify the bicycle facilities, similarly to what done by Landis [142]. By taking inspiration from the Dutch bicycle infrastructure design guidelines, and considering factors such as speed limit, lane width, presence of parking facilities, they propose a novel classification for allocating levels of traffic stress across the four cyclist types (see Fig.2.3). They created the LTS 2 category, which aligns with the 'interested but concerned' demographic [64, 160]. The evaluation of intersections also takes into account their dimensions and the speed at which crossings occur [38]. LTS 1 is tailored to the 'children' cyclist type, aiming to provide a comfortable riding experience for users of all skill levels, necessitating significant separation from motorized traffic. LTS 3 caters to the 'enthused and confident' cyclists, who are more willing to navigate higher traffic stress levels. Finally, LTS 4 is designed for the 'strong and fearless' group, characterized by mixed traffic conditions, high vehicle speeds, and substantial traffic flow. The determination of a path's stress level is based on the highest stress level found among its constituent elements (also known as weakest link logic), which include streets, crossings, or intersection approaches. To be more clear, two nodes are considered to be linked to a specific LTS when there's an edge between them with a peak LTS of a certain value. Interestingly, in their study, they found for San Jose case (California, US) that routes with low stress results in the formation of numerous isolated clusters, also known as 'islands'. These clusters are segregated from one another by physical barriers like highways, rivers, and major arterial roads without low-stress crossings (the so called high-stress links) [160]. Overall, this method aims to assess network connectivity based on five dimensions, which are directness, coherence, route complexity, safety and comfort. They demonstrated that improving low-stress network connectivity can lead to a considerable increase in the number of people who choose to bicycle for transportation. Moreover, they highlighted the importance of addressing gaps and barriers in the existing bicycle network to ensure continuous and low-stress routes for cyclists [160].



Figure 2.3: Levels of Traffic Stress. Adapted from [Alta Planning](#)

2.5.2 LTS research and Bicycle Network Connectivity

In contemporary transportation research, the Levels of Traffic Stress (LTS) framework is emerging as one of the predominant models, owing to its applicability across various domains and to its flexibility and use of OSM and public administration data. Additionally, as it will be here discussed, the LTS model is instrumental in guiding the planning and development of new cycling infrastructure. Here we discussed three pivotal aspects such as the strategic enhancement of bicycle networks, the impact of LTS on cycling dynamics and the importance of centrality measures within bicycle networks and LTS analysis. Particularly, the application of centrality metrics to evaluate flow within bicycle networks marks a significant trend, emphasizing the enhancement of cyclist safety and accessibility – core objectives in the LTS methodology. This novel utilization signifies an expansion of network theory principles, specifically adapted for the assessment and improvement of cycling infrastructure. Regarding LTS application, Lowry, Furth, and Hadden-Loh [150] computed bicycle stress using Marginal Rates of Substitution (MRS) for every road segment concerning stress-inducing and stress-mitigating factors. The goal of his work was to present a new approach to prioritize bicycle improvement projects based on accessibility of key destinations. MRS values, which factor in the number of lanes and speed limit of a street, serve as input parameters. The analysis pinpoints specific neighborhoods or areas of Seattle that are expected to have greater bikeability scores. Several other cities and towns across the United States, including Seattle, are actively improving their cycling infrastructure as proved by the work of Murphy and Owen [171] focused also on Washington D.C., Minneapolis and Miami. Here, the Bicycle Level of Traffic Stress metric has been modified to assign stress levels to roads and intersections based on OpenStreetMap data, and job accessibility calculations are carried out for the resulting low-stress bike networks. Subsequently, a measure termed 'access gap' was developed to assess the disparity between low-stress and higher-stress access within urban cycling networks. This metric aims to identify areas for potential enhancement by upgrading high-stress routes. The research revealed that restricting cycling routes to only low-stress networks consistently diminishes job accessibility. The degree of this decline varies not only between different metropolitan regions but also within distinct neighborhoods of the same city. A similar analysis focused on jobs accessibility was done by Imani, Miller, and Saxe [124] for Toronto. Specifically, they categorized roads and intersections using LTS and analyzed the 30-minute cycling isochrones for job and population access across different LTS levels. They also found that lower LTS levels correlate with reduced job accessibility by cycling. However, to be more precise, the majority of LTS implementations in the literature have been adjusted to accommodate the specific characteristics of their originating locations [150]. Another example in this sense is what done by Montgomery County Planning Department [166], which altered

the initial LTS classification and introduced LTS 0, LTS 2.5, and LTS 5 categories to suit Montgomery County, Maryland's specific requirements. This modified methodology considers various factors such as roadway width, speed limits, intersection design and separation from motorized traffic, which were not fully addressed in the original method. LTS-0 is a new category of bikeway for completely separated bicycling infrastructure, while LTS-2.5 creates a new category because the gulf between the LTS-2 and LTS-3 is large, so an intermediate step is needed. This is often used as a minimum for industrial streets with high traffic volume. The creation of the fifth level is less useful from a cyclist point of view, since it has very high traffic speed. Furthermore, Moran et al. [168] introduced a large-scale approach to low-stress bicycle connectivity analysis by analyzing the potential connectivity improvements on an individual street segment basis. Scripts and database tools were utilized to assign traffic stress levels to the road network. This approach aimed to explore whether targeted interventions could foster low-stress connectivity. The study focused on Philadelphia, where researchers navigated through various low-stress 'islands' and calculated the shortest paths among millions of origin-destination pairings. This process helped identify specific road segments that would most benefit from such interventions. Ultimately, a prioritized list of links connecting these islands was generated. The findings indicate that the primary routes were interconnected through roads with a LTS of 3, highlighting these as the critical links requiring enhancement. However, as previously mentioned, not all factors that can influence an individual's choice and risk perception are accounted for by the LTS. For instance, we can cite the fact that Chen et al. [51] analyzed the relationship between accidents and LTS levels. By looking at the crashes and injuries severity, a correlation was done with LTS levels. They found that the more severe ones are more likely to happen in LTS 4 compared to LTS 3. Moreover, Cervero, Denman, and Jin [49] work based in Europe tried to correlate LTS values with work commute data across 36 UK cities. The study showed that the share of bicycle usage rises with the presence of lower stress roads (LTS 1 and LTS 2). According to their model, there is a 0.73% increase in bicycle usage for every 10% increase in low-stress links. This is significant considering the average cycling rate of 6.2% observed in the case studies analyzed. Other LTS implementations have been developed to overcome place-specificity and generate a more generalized LTS. In LTS second version proposed by the original authors [90] the goal was to enhance the LTS's generalizability by employing only six of the original 21 variables. Eldred [79] and the all Conveyal team presented an extreme simplification of the traditional LTS, termed 'Surrogate LTS' utilizing only four variables derived from a few OpenStreetMap (OSM) tags. Here, a small set of relevant variables commonly used in previous methodologies was used. These included road width, number of lanes, presence of cycling infrastructure, presence of heavy vehicles, traffic speed, and traffic volume. While this initial approach might not capture every detail of bicycle connectivity, it enables LTS analysis when specialized data is unavailable. The proliferation of LTS adaptations addressing both place-specificity and generality requirements leads to varying LTS classifications across the same road networks. In the current state of the literature, it is important to note that environmental factors such as road gradient are only partially considered and not included as a factor, as was hoped for in Mekuria, Furth, and Nixon [160] original work, within the LTS classification. For example, in the work of Delaware Valley Regional Planning Commission [71], the addition of the slope as an external factor wasn't included inside the classification but only in the connectivity analysis together with others with specific focus on schools' connection to find the shortest and less stressful paths. Specifically, these factors were included in the 'overall cost' calculation for each road segment, which also takes into account the length and assigned LTS (Cost = Link Length \times (1 + Link LTS + Link Slope Factor)). Interestingly, Halefom et al. [107] selected the average traffic stress level of preferred routes as their study's dependent variable. Their findings revealed an inverse relationship with the average elevation change per segment. In other words, to select a less stressful path,

individuals, especially women and those riding for leisure, usually opt for routes with steeper inclines to circumvent more direct routes that have higher traffic stress, typically found in flat terrain areas. Recently, Furth, Sadeghinassr, and Miranda-Moreno [91] divided the concept of comfort in high traffic stress and avoidance of steep climbs. The study was done in Montreal, an area that does not feature any particular mountainous relief, except for a hill called Mont Royal, which is 233 meters high. Authors proposed a 'Steepness Criteria' that can complement Level of Traffic Stress in order to provide more realistic results in term of accessibility. By applying it on Montreal, it was found that imposing a Steepness Level of 5.0 (maximum slope of 5%) only affects the accessibility of a few neighborhoods. However, when a stricter Steepness Level of 3.5 (maximum slope of 3.5%) is applied, many more neighborhoods lose accessibility. This significant disparity highlights the context-sensitive nature of such impacts, suggesting that in some areas, the effect of a 5.0% steepness limit might be much more substantial. Regarding centrality measures and bicycle connectivity, these measures are commonly used to estimate flow on networks [237]. Specifically as highlighted by Vybornova et al. [237] in its literature review, authors like Zhang et al. [252] developed a statistical model linking edge betweenness centrality metrics with data on cyclist related traffic accidents. Ye, Wu, and Fan [248] predict traffic flow by using betweenness centrality and merged travel demand information. Finally, Kamel and Sayed [130] investigated how node and edge betweenness centralities in the bicycle network correlate with the number of traffic accidents and kilometers cycled. They found that these measures serve as indicators of the overall quality of the bicycle network. However, network measures, like edge betweenness centrality, are found to be dependent on changes in network boundaries, an effect called 'border effect' [194]. Several strategies should be adopted to smooth this effect. For example, by introducing a cut-off radius for the edge betweenness centralities calculation [247].

2.5.3 Gap Analysis

A key goal for a city's sustainability is to bridge the discontinuities in cycling paths, thereby forming a more interconnected and extensive biking network [102]. However, the expansion of bicycle infrastructure often faces challenges due to political stagnation rooted in the complex issue of car dependency [155]. As an example, over the course of a century, Copenhagen underwent extensive political efforts to establish a protected on-street bicycle infrastructure. However, 300 separate sections are still present [173, 237]. Research has established a correlation between the prevalence of bicycling and the availability of corresponding infrastructure [39]. A study by Handy and Xing [111] shows that bicycle commuters are notably distance-sensitive, so this can influence their choice to use specialized facilities. Hence, it's pivotal to construct a network offering direct routes with minimal deviations. Mekuria, Furth, and Nixon [160] emphasize the significance of connectivity in *low stress* networks by assessing route quality based on their most challenging sections that may be unsuitable for those with a lower tolerance. According to Schoner and Levinson [213], the effectiveness of dedicated infrastructure is strongly tied to the degree of connectivity it offers. Planning isolated segments without integrating them into the larger street network diminishes the overall usefulness of this infrastructure. Linear regression analyses performed by Schoner and Levinson [213] have shown that the factors of connectivity and directness play significant roles in forecasting bicycle infrastructure use. According to Szell et al. [227] and Vierø, Vybornova, and Szell [235], traditional random growth strategies for bike infrastructure are not useful, because they can delay the development of functional cycling infrastructure and potentially lead to criticisms about underused bike tracks. Instead, a more systematic approach is recommended. The focus should be on the concept of directness. Ilie et al. [123] use the previously mentioned the Bicycle Level of Service (BLOS) procedure to a new planned bicycle network and categorized the street segments with the lowest scores as the gaps in

the network. However, according to Vybornova et al. [237], efforts to accurately identify the most critical gaps or 'missing links' in a city's network have been scarce. The author notes that their work falls into a category of studies aimed at developing scalable strategies for bicycle planning, originating from the analysis of a single city, in this case Copenhagen and '*repairing*' the network [144, 204, 250]. The IPDC (Identify, Prioritize, Decenter and Classify) procedure formalized by Vybornova et al. [237], is structured in this way: the first step, which is Identify, consist in taking OpenStreetMap data to map an urban network of streets and protected bicycle tracks. Data has 'unprotected' links, which are street segments for motor vehicles without protected bicycle infrastructure and 'protected' links that represent areas with protected bicycle infrastructure. On the other hand, nodes connected only to one link type are labeled as either 'protected' or 'unprotected', while nodes with both link types are 'contact nodes'. By using shortest paths (Dijkstra all-pair shortest path algorithm) between two contact nodes and discarding parallel ones with a specific detour factor, gaps are then found. A gap can be intended as a continuous segment of missing protected infrastructure. The positive impact of the approach described is aimed at reducing the distance cyclists have to share the road with motorized traffic, thereby decreasing their stress. The goal is to focus on closing gaps located on the most frequented bicycle routes. Then, for the second step, the authors propose assessing the priority based on centrality, cost, and benefit. Centrality involves determining how crucial the missing link is within the network. Cost refers to the expense involved in closing the gap. Then the benefit is measured by how many people will gain from the gap's closure, quantified by the reduction in meters cyclists have to travel alongside motor vehicles. To do this and identify the most used bicycle routes and thus the most critical gaps, the authors use a metric termed '*link betweenness centrality weighted by gap length*'. The betweenness centrality of a link shows its frequency of use in these shortest paths across all node pairs in the network. By weighting centrality with the gap's length, they estimate the total meters cycled in mixed traffic for each gap. The assumption here is the fact that for every origin-destination pair, a cyclist takes the shortest route, allowing calculation of the proportion of cyclists expected on each link. To mitigate the 'network edge effect' authors factor in a locality parameter (λ) which skews centrality metrics towards the network's center. Lastly, a calculation of the 'gap closure benefit' is done for each gap, combining the betweenness centrality with the gap's length. Dividing this by the gap's total length yields a measure of expected meters cycled per unit of investment, helping prioritize which gaps to close. Regarding the declustering step, this is performed because in many cases prioritized gaps in bicycle networks overlap and form complex clusters instead of simple paths. Gap clusters often emerge because each network node is a potential origin or destination, gaps are composed of car links and start/end at contact nodes and networks typically have high node density at multi-lane intersections. The declustering heuristic used to simplify gap clusters use the same benefit metric. Finally the last phase is the one called classification. Basically, with the remaining gaps, the classification system, developed through hands-on inspections and field visits in Copenhagen, divided the gaps in different categories, namely Street (ST), Intersection (IS), Right-Turn Lane (RT), Bridge (BR), Roundabout (RA), and Error (ER). This scheme aimed to simplify the analysis of bicycle network results and assist in the decision-making process for future planning. Interestingly, it was found that by comparing results with the city's most recent Cycle Path Prioritization Plan, several overlaps occurred.

2.6 Bicycle Network Analysis Score

The Bicycle Network Analysis (BNA) is a methodological approach created by the PeopleFor-Bikes [189] organization, designed to quantify the effectiveness of bicycle networks within urban

environments. Specifically, its goal is measuring how well the bicycle network in a city connects the citizens with the destinations. It transcends traditional metrics like the total length of bike lanes or proximity to bicycle facilities, which, although somewhat indicative of a city's bike-friendliness, do not fully capture the essence of a seamlessly integrated network of cycling routes and the significance of destinations in creating a cohesive biking infrastructure. Specifically tailored for communities within the United States (first started with 300 cities) but also extended to other continents, this approach forms a core component of the Places For Bikes City Ratings initiative [10, 189]. The core component of the BNA score is the Level of Traffic Stress (LTS) classification developed by Mekuria, Furth, and Nixon [160] which is previously analysed in the specific section. To summarize, the LTS framework categorizes road infrastructure conditions based on the level of stress they impose on different types of cyclists. Factors influencing this stress include traffic speed, street type and intersection design. The original publication highlighted the disproportionate impact of high-stress intersections on route continuity and the *weakest link principle*, implying that the most stressful segment of a route disproportionately defines its overall stress level. It is based on a block to block analysis that involves associating roads with census blocks (or analogous units), calculating the shortest path between blocks, both unrestricted and restricted to low-stress segments and finally compare these paths to ascertain viable low-stress connections. An important part of the process involves the evaluation of the total number of destinations (universities, schools, supermarkets, doctors, bar etc.) around each block, of the subset reachable via low-stress paths only and the comparison of these sets to derive a score reflecting accessible destinations. The score ranges between 0 and 100, which is assigned to each census block and then aggregated to the whole city or town. The methodology is based almost entirely on OSM data and this provides reproducible results. Moreover, other datasets are often used (for example: data regarding population distribution for each census block or jobs distribution). Overall, the BNA's block-based approach and its emphasis on low-stress connectivity provide a nuanced understanding of a bicycle network's effectiveness. Destination category provides flexibility thanks to its ability to assign weights to different destination types in the final scoring. In the context of Italy, the cities that were notably emphasized include: Bologna (BNA score 73), Milan (BNA score 69), Verona (BNA score 66), Roma (BNA score 50) and Napoli (BNA score 49) [190].

2.7 Community-Based Mapping: The OpenStreetMap Phenomenon

OpenStreetMap (OSM) is a collaborative initiative aimed at developing a comprehensive, freely editable map of the globe. This project is driven by a global community of contributors who gather and refine data pertaining to a multitude of geographical elements, including but not limited to transportation routes, topographical features and points of interest [98].

2.7.1 Formal Description

OpenStreetMap (OSM) is an extensive, user-generated digital mapping platform, often likened to the 'Wikipedia of maps' born in 2004 at University of London [183]. It represents a comprehensive, editable map of the globe, compiled through the collaborative efforts of a vast community of mappers. These contributors systematically gather and refine data about a multitude of geographical features and points of interest across the world. OSM's ethos is grounded in the principles of open data and communal contribution, facilitating a dynamic and accessible mapping resource [167]. Its popularity has only shown an exponential increase along the years. By

early December 2023, over 11 million users were registered on the platform [182]. However, by December 2023, almost 3 million only accumulated contributors per month were actively using the platform.

2.7.2 OSM Data

The OSM project harnesses a variety of methods for data acquisition, including manual surveying using GPS devices, digital imaging techniques and the utilization of other open-source data. This expansive database encompasses a wide spectrum of geographical elements, ranging from intricate road networks and pathways to physical landmarks. Additionally, it covers detailed aspects of the built environment, including but not limited to, infrastructure facilities, public amenities, and points of interest such as dining establishments, accommodations and cultural sites [167, 215]. The platform is equipped with a diverse array of tools designed to accommodate varying degrees of mapping proficiency. For routine editing tasks, the widely used iD Editor, a web-based interface, is typically preferred. Conversely, for users with more advanced skills, JOSM (Java OpenStreetMap Editor) offers a sophisticated desktop alternative but the list of other alternatives is very long [106]. At the core of OpenStreetMap's data architecture are three primary elements:

- nodes: defining points
- ways: linear features constituted by nodes
- relations: intricate structures composed of amalgamation of nodes, ways and other relations if available.

2.7.3 Tagging Process and Limitations

The tagging system in OSM is a crucial aspect where community members assign labels or 'tags' to various map elements to denote their characteristics. This process is integral to categorizing and detailing the diverse components within the map [106, 167]. A tag in OSM is composed of two parts: a key and a value. The key denotes the category or type of the feature and the value provides specific information about that feature. For example, a tag with the key 'highway' and the value 'residential' indicates that a particular way is a residential road. This system is highly versatile, allowing for a wide range of attributes to be assigned to map features [167]. Tags within OpenStreetMap are designed to encapsulate a range of attributes, extending from tangible physical characteristics to more intricate and specific details. For instance, they can delineate aspects of a roadway such as its surface composition, the count of traffic lanes it comprises or even the designated nomenclature of various points of interest. In addition, these tags are capable of conveying nuanced information such as wheelchair accessibility [106]. The structure followed for tags is hierarchical, where more general categories are broken down into more specific details. For example, under the broad category of 'amenity', one might find more specific tags like 'restaurant', 'bank', or 'school' [67]. As mentioned above, the governance and continuous development of the map heavily depend on the active participation of its global community. This involvement is not without its challenges. Issues such as data availability, quality and consistency [115, 251] occur. This is especially noticeable in areas that are less populated or not as well-known. In these regions, the amount of available data is significantly less compared to the abundant data found in densely populated urban areas [17]. Herfort et al. [115] discovered in their spatial analysis that in 69% of the studied cities, representing 48% of the world's urban population, the completeness of building footprint data in OpenStreetMap (OSM) did not exceed 20%. On average, the completeness of urban OSM building data globally was a

mere 24% per urban center. This was particularly evident in regions such as East Asia & Pacific, the Middle East & North Africa and South Asia. The OSM data is licensed under the Open Database License (ODbL), which permits free usage, modification, and sharing, provided there is appropriate attribution to OSM and its contributors and to ensure that any products derived from this data are similarly shared freely under the terms of the Open Data Commons Open Database License [181]. This open licensing model underpins applications in various domains, from navigation systems to humanitarian initiatives, urban planning and geospatial research [129].

2.8 Case Studies

I will provide a concise overview of the two case studies relevant to this study.

2.8.1 Trento

This study focuses on the city of Trento as a key case study for analyzing bicycle levels of traffic stress, safety and accessibility. Trento, a municipality in Northern Italy, is situated in the Adige River valley in the Trentino-Alto Adige/Südtirol region. The city is enveloped by the majestic peaks of the Dolomites, a section of the Southern Limestone Alps, which provides a picturesque backdrop. Geographically, Trento is notable for its diverse landscape. The topography of Trento includes steep mountain slopes and rolling hills and expansive valleys, which are intersected by the Adige River and its tributaries. The municipality covers an area of approximately 157 square kilometers, with altitudes ranging from about 55 meters in the valley floor to over 2000 meters in the mountainous areas. The official altitude is 194 meters [60]. As of December 18th 2023, Trento recorded a population of 118291, with a population density of 749.25 ab./km² [126] and spanning a total area of 157.88km² [230]. Trento has developed urban infrastructure that reflects a balance between vehicular and non-vehicular transportation. According to Comune di Trento [59] it primarily extends along a north-south axis, featuring a core area of significant value, the historic center, which is protected by a limited traffic zone. Along this axis runs the city's main thoroughfare, the Trento ring road, designated for the distribution of vehicular traffic. This ring road, approximately 15 kilometers long, largely consists of two lanes in each direction on separate carriageways. It starts from the south, where it connects to the A22 Brennero highway coming from Verona, and extends towards the northeast, where it joins the SS47 Valsugana highway. In the article *la Top Ten delle città italiane più bike friendly: l'Emilia Romagna primeggia* [141], it is pointed out that the city's layout, characterized by a blend of modern and historic architecture, provides ample opportunities for cycling. This includes well-designed bike paths, pedestrian-friendly zones and an abundance of natural scenery, with parks and green spaces constituting a significant portion of the city's area. According to the authors, in 2019 it offered 0.47 meters of bike paths per inhabitant and a total of 55.12 km of bike paths. The same article ranked in 2019 Trento as the ninth most bike-friendly city in Italy [141]. Moreover, in terms of sustainable mobility, that is, through the use of bicycles, pedal-assisted bikes and scooters, it ranks fifth in Italy [96]. In 2023, ComuniCiclabili, an initiative by FIAB, acknowledged and celebrated the Local Administration of Trento for its dedicated implementation of substantial policies aimed at enhancing bicycle mobility. This recognition underscores Trento's commitment to promoting bikeability within the city. As a testament to these efforts, FIAB renewed the assignment of the prestigious yellow flag to Trento, signifying its continued excellence in fostering a bicycle-friendly environment [61]. Although the current situation isn't dire, there's still considerable room for improvement in Trento's cycling infrastructure, as underscored in a recent article [54]. According to the author, even with recent interventions and new fifteen bike routes, the assumed cycling

goals will not be achieved. Specifically, there is a near absence of continuity in the various cycling routes traversing the city. According to the analysis of the 75 km of cycling routes cataloged in the Biciplan [59] only about 25 km constitute true cycling lanes with a separation from pedestrian paths. This separation is predominantly achieved through a white demarcation line, as these paths are parallel and contiguous to the sidewalk without any interruption. The remaining ones, are made up of mixed-use pedestrian and cycling sidewalks, with the result of several interferences. Sometimes bus stops are placed directly on the cycling path with the result of dangerous interferences. In contrast, the urban planning approach in Bolzano exhibits a distinct paradigm where bicycles seldom coexist spatially with pedestrian traffic. In this model, the bicycle pathways are frequently aligned with vehicular lanes, maintaining a similar elevation. Crucially, these cycling routes are safeguarded from vehicular traffic by either a verdant buffer zone or a raised curb, serving as a physical barrier to enhance safety and delineate the distinct zones of transportation. Moreover, if not well-maintained, horizontal signage fades within a few months, making it difficult to distinguish the designated areas for pedestrians and cyclists. These identified issues significantly compromise the viability of the initiated proposals and projects, necessitating thorough analytical examination.

2.8.2 Bolzano

This study centers also on the city of Bolzano as a case for evaluating bicycle traffic stress, safety and accessibility. Bolzano, a city in Northern Italy, is located in the heart of the South Tyrol region, nestled in a basin amidst the confluence of the Talvera, Isarco and Adige rivers. It is surrounded by the peaks of the Dolomites and the Alps, offering a stunning natural setting. Geographically, Bolzano is characterized by its varied landscape, which includes a mix of flat valley floors, gentle hills and rugged mountain terrain. Thanks to its strategic position, Bolzano serves as a key transportation and logistics hub, connecting Italy with northern Europe. The city covers an area of about 52.3 km², with elevations ranging from around 262 meters in the lower urban areas to higher altitudes in the surrounding mountains. As of December 28th 2023, Bolzano recorded a population of 105939, with a population density of 2025,99 ab ab./km² [127] and spanning a total area of 52,29 km² [29]. The urban area extends less spread and more homogeneous compared to Trento. Regarding the cycling facilities, the entire municipality is interlaced with a network of several bicycle paths spanning a total of 76 kilometers (0.71 meters of bike paths per inhabitant), making it the fourth Italian city with the highest number of bicycle commutes [29, 141, 218]. Also in absolute terms of sustainable mobility numbers, it ranks fourth in 2021 [96]. ComuniCiclabili has recognized and commended the Bolzano Local Administration too for its steadfast commitment to enacting significant policies that improve bicycle mobility and it has reaffirmed Bolzano's status as a premier cycling-friendly city by re-awarding it the esteemed yellow flag, symbolizing its ongoing excellence in promoting a conducive environment for bicycling [61]. In this case study, we project encountering outcomes that are comparatively less pronounced than those in Trento, attributed to Bolzano's more advanced status in cycling infrastructure and accessibility. Nonetheless, we still expect to reveal significant insights into traffic stress. This is based on recent studies that have documented a 30% increase in automobile usage among Bolzano's residents, a trend that persists despite the city's substantial bicycle infrastructure [5].

Chapter 3

Methodology

This chapter provides a comprehensive overview of the methodology adopted for the research, beginning with the meticulous selection of data sources, the variables involved, and the careful choice of software necessary for the study's tool implementation and execution. It then delves into an in-depth explanation of the complete data analysis journey, encompassing a diverse range of analytical approaches. These include a detailed investigation into traffic stress levels, an exploratory examination of spatial data, and comprehensive network analyses. The chapter further extends into intricate cluster and gap analyses, alongside a thorough investigation of accident patterns. Additionally, a unique element of this study is the development and application of a cycle network analysis score, specifically tailored to assess cycling safety and efficiency. The chapter then transitions to describe the rigorous testing phase conducted in Trento, complemented by a validation process in Bolzano. This dual-phase approach ensures the robustness and applicability of the findings. The culmination of this chapter is the section on Final Predictive Analysis. This segment synthesizes the insights garnered from the preceding analyses to construct predictive models. These models aim to identify high-risk areas within the road network, drawing a correlation between areas of high traffic stress and the frequency of accidents. This predictive framework is instrumental in proposing targeted interventions for enhancing road safety and efficacy.

3.1 Data Sources

For this study, various open-data sources were utilized. A significant portion of the preparatory work involved developing a system to manage these diverse data types in a consistent manner. In terms of broad categories, the majority of the geospatial data used were vector data, while a smaller portion consisted of raster data. Vector data refers to a type of spatial data used in Geographic Information Systems (GIS) and various mapping applications like OpenStreetMap (OSM) [180], Google Maps [99], etc. It represents geographic features through points, lines, and polygons. Each shape is defined by X and Y coordinates and may include additional data attributes. Specifically, *points* represent specific locations such as cities or landmarks; *lines* illustrate linear features like roads and rivers; and *polygons* delineate areas including buildings or lakes. Vector data is notably precise in depicting boundaries and linear features and it efficiently represents complex features with detailed attributes. In contrast, raster data is a grid of cells or pixels, where each pixel has a value representing information such as temperature, elevation or spectral data. This type of data is commonly used for digital images, aerial photographs and satellite imagery. The resolution of raster data is determined by the size of the cells, with

smaller cells providing higher resolution. These data types were chosen for their complementary differences and their utility in various analytical approaches. Vector data is particularly valuable for mapping and analyzing specific locations, paths and networks, making it highly useful in urban planning, mapping and transportation. Raster data, on the other hand, is more commonly employed in environmental studies to address gradual spatial changes. While vector data excels in network analysis, raster data is optimal for surface analysis. However, as will be discussed later, these two types of analysis are often used together, as each can provide valuable insights to enhance the understanding provided by the other [52, 169]. In the selected case studies, we primarily utilized OpenStreetMap (OSM) data for gathering detailed information on street routes accessible to bicycles, including bicycle lanes, express lanes, and mixed routes. Additionally, OSM served as a vital resource for acquiring data on buildings within the study areas, particularly for identifying and extracting information on buildings categorized as 'Common Destinations' or Points of Interest. This information was crucial for our subsequent analyses. Our preference for OpenStreetMap over Google Maps is rooted in its status as a prominent example of an open-data geographic information project. OpenStreetMap is renowned for providing a freely accessible, comprehensive tool that spans the entire globe. Its data, which can be updated or modified by users, offers a unique level of accessibility and community involvement. However, this approach is not without its drawbacks. One significant limitation is the variability in data availability [115], which is particularly pronounced in less populated or renowned areas. In such regions, the volume of available data is markedly lower compared to more densely populated urban areas [17]. Specifically, Herfort et al. [115] found that, for example, in 69% of the analyzed cities, which are home to 48% of the global urban population, OSM building footprint data did not reach 20% completeness. The global average urban OSM building completeness was found to be only 24% per urban center. This incompleteness was especially pronounced in areas like East Asia & Pacific, the Middle East & North Africa, and South Asia. Furthermore, the data quality and consistency across different states varies considerably [115]. While there are established guidelines for tagging within the OpenStreetMap community, adherence to these guidelines is not uniformly strict. As highlighted by Zhang and Pfoser [251], due to the lack of quality control in data collection, errors in OSM data can be significant. Earlier research indicated that the error in OSM data compared to British Ordnance Survey datasets was within 6 meters, with a 26% coverage. This inconsistency necessitates additional effort from users in analyzing and interpreting the data. These mentioned challenges highlight the need for a careful and nuanced approach when employing OpenStreetMap data in geographical analyses. This section showcases a selection of OpenStreetMap tags meticulously extracted during the initial stage of our analysis as it can be seen from the Table 3.1.

Data Type	Description
Buildings	Extracted using the tag <code>{'building': True}</code>
Road Network	Types: primary, primary_link, secondary, secondary_link, tertiary, tertiary_link, trunk, trunk_link, residential, cycleway, living_street, unclassified, motorway, motorway_link, pedestrian, steps, track

Table 3.1: Summary of OSM Data Downloaded for the different analysis

Moreover, since the Level of Traffic Stress Analysis required the calculation of the slope of the streets, raster data were then needed for this purpose. To do this we used high resolution elevation data (10 m grid) from Tinitaly [228]. The Data Elevation Model (DEM) provided is encoded as 'ESRI ASCII Raster' derived from the original DEM in a Triangular Irregular Network (TIN) format. The database consists of 193 square tiles, each 50 km on a side, and is designed

to provide a seamless representation of Italy's topography. To download data related to the area of interest, there are various services available including WMS (Web Map Service), WCS (Web Coverage Service) and WMTS (Web Map Tile Service). The Web Map Service (WMS) provides images of geospatial data, like maps, in various formats. The Web Coverage Service (WCS) allows for the retrieval of geospatial data as 'coverages' - digital representations of geophysical or geographical data. The Web Map Tile Service (WMTS) serves map tiles, which are pre-rendered, fixed-size images that make up a larger map. Each service offers different ways to access and visualize geospatial information, depending on the user's needs. It is also possible to manually download the areas of interest as Tagged Image File Format (TIFF) by simply clicking on the squares of interest and merge them later using other tools [228]. A TIFF file is a versatile format utilized for storing high-quality graphics and images. Owing to its minimal compression, these files can be notably large. However, they uniquely support the storage of multiple image layers, making them especially valuable in geomorphological analyses (Figure 3.1).

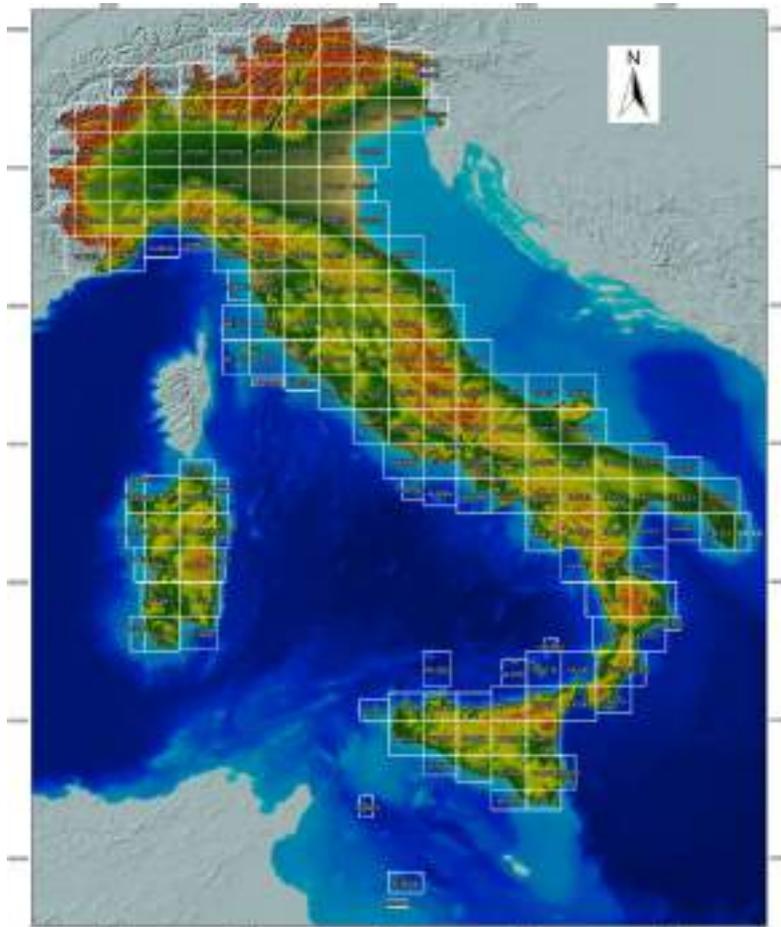


Figure 3.1: Interactive Grid of Italy's Downloadable Area: A Detailed View of the Geospatial Selection Interface

In addition to these two types of data, for the analysis related to the BNA score, it was necessary to obtain population distribution data within the area of interest to achieve more

accurate and less approximate data. In doing so, we used a geodataset from *Kontur Population Italy Dataset* [136] that consisted of an H3 grid with 400 meters resolution. This is provided by the Humanitarian Data Exchange, which is an open platform for sharing data across crises and organizations. It aims to make humanitarian data easy to find and use for analysis, designed to facilitate access to data for humanitarian operations. The platform hosts datasets from various sources and covers a wide range of humanitarian issues, enabling more informed decision-making and response planning in crisis situations [120]. The H3 grid with the 400 m resolution used is a geospatial hierarchical indexing system that partitions the world into hexagonal cells (Figure 3.2). It's used for various applications, such as mapping and spatial analysis. This grid system, developed by Uber, offers several advantages like uniform cell size, no singularities at the poles, and consistent spatial resolution [105, 104].

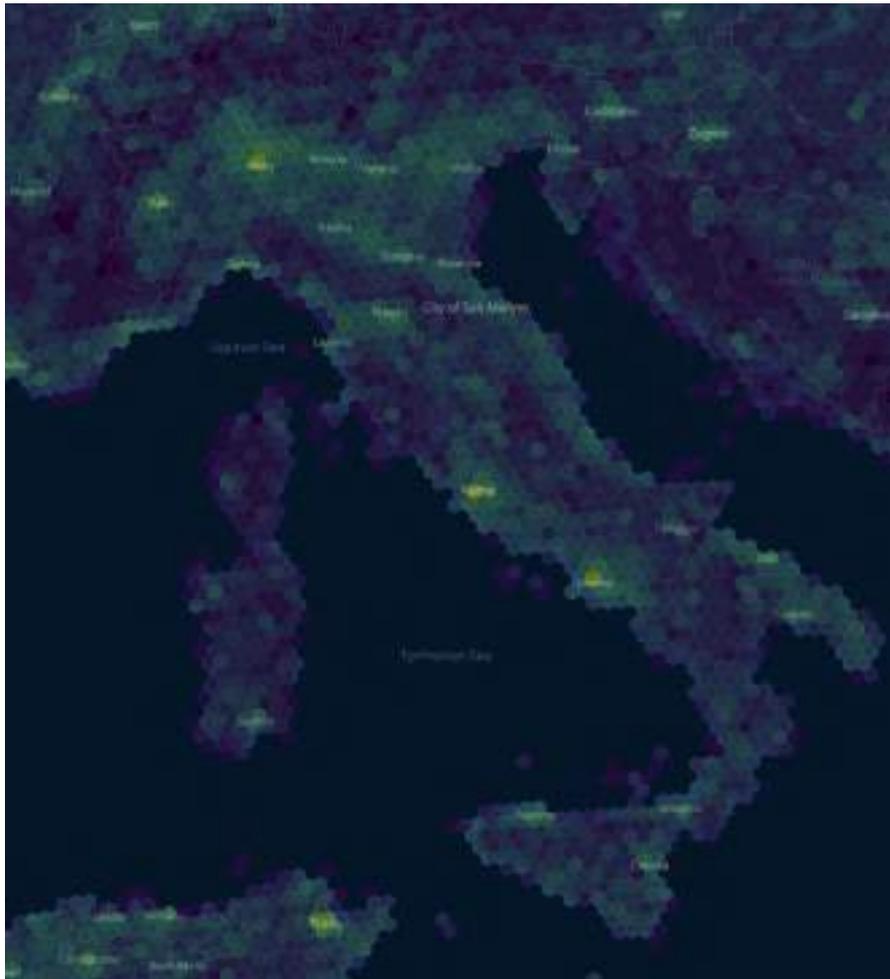


Figure 3.2: Population Density of Italy Visualized in 400m H3 Hexagons by Kontur, Courtesy of Humanitarian Data Exchange

In the analysis of accident data within the designated road network, methodologies varied based on the reference network. For Trento, generic accident data were obtained from the regional open data portal, encompassing records maintained by the municipal police since 2005 [125]. These monthly updated datasets, while lacking in detailed classification of incidents,

align geographic coordinates with accident occurrences and dates. Analytical focus was confined to the most recent five-year period to ensure relevance to current conditions (Figure 3.3). In contrast, Bolzano's data, potentially involving bicycle-related incidents, was sourced through the WFS service with tailored data filtering [241]. Detailed methodologies for this case study are elaborated in a subsequent section.



Figure 3.3: Close-Up View of Traffic Accidents in the City of Trento Over the Last 5 Years

3.2 Tool Design, Execution and Software

3.2.1 Tool Design

As highlighted in the previous section, various methodologies exist for analyzing bicycle paths, focusing on both their infrastructure and qualitative aspects, as well as their accessibility to popular destinations. Some approaches concentrate primarily on the Level of Traffic Stress [118, 150, 160, 203], others exclusively on the Bicycle Network Analysis (BNA) score [1, 2, 191], and yet others on the structural aspects of the network like centrality measures [156, 176, 237]. This work aims to provide public administrations and other interested parties with a comprehensive

tool that integrates both infrastructural and qualitative data, offering insights closely aligned with the realities of the Italian context under study. This initiative also arises from the need to address the frequent lack of complete data [115]. A particular emphasis is placed on safety towards the end of the analysis, including road accident data and their correlation with previously examined aspects. This approach aims to meet the requirements of Sustainable Urban Mobility Plans (SUMP) and Biciplan in terms of safety for cyclists [84, 225]. The idea consists in developing a suite of scripts, primarily executed locally and fundamentally based on Python, R and their associated libraries. Figure 3.4 delineates the four major processes into which the scripts are categorized, serving as a generic representation of our tool. These primarily consist of Python files or Jupyter notebooks with Python code, which are designed for individual analyses or report generation, with integrated communication for data and function sharing as needed. The R component is specifically utilized for slope calculations using raster data, selected for its libraries that offer enhanced computational capabilities (e.g., the *slopes* library), which will be discussed in a dedicated section. All scripts are housed within the *code* directory. Generated data and graphics are stored in the *data* and *images* folders, respectively. Particularly, the *images* folder contains subfolders named after the analyzed areas, ensuring organized storage of images. This system aids in distinguishing between various outputs. Specific files related to areas of interest within the *data* folder are named using an abbreviated format of the area's name (e.g. 'Trento_all_lts.csv'). Given the tool's specific objective to validate analyses in merely two case studies, it does not incorporate methodologies for optimization or automation through cloud computing services (e.g. AWS or GitHub Actions). Similarly, to ensure a streamlined design with minimal technical dependencies, the integration of complex database systems like PostgreSQL was deliberately omitted. Nevertheless, this tool possesses inherent adaptability, rendering it suitable for modification to accommodate analyses of larger geographical regions with voluminous datasets. This scalability feature enables the tool to potentially act as a foundational tool for more extensive and intricate future analytical endeavors. The tool includes a Python file (*main.py*) which functions as an orchestrator, interlinking all the files within the directory. Here, the core function, *get_autocomplete_suggestions*, is implemented to query the Nominatim service provided by OpenStreetMap. This service facilitates the acquisition of geospatial data based on user input. Utilizing a GET request, the function sends a query to the Nominatim API, with parameters structured to return JSON-formatted responses limited to five suggestions. This limit is set to streamline the suggestion process. The function then parses the JSON response, extracting and returning a list of suggested place names, provided the response status is successful (HTTP 200). Following the acquisition of suggestions, the user is prompted to confirm or correct their input, ensuring the precision of the geographic data to be processed. Subsequent to user confirmation, the script executes system commands to initiate further data processing. This involves running additional Python scripts and Jupyter notebooks, which are part of a larger data analysis pipeline. The *os.system* calls are used to automate the execution of other scripts and Jupyter notebooks. In summary, this setup ensures that the user, after duplicating the main repository, upon installing the necessary dependencies and packages, merely needs to input the area of interest in the console. Following this simple action, the tool automatically executes all relevant files, streamlining the analysis process. This design choice significantly simplifies the user interaction with the tool, focusing on ease of use and efficiency in handling complex tasks. In the conclusion of our tool, we have a Python script designed to generate a comprehensive HTML report. This report effectively consolidates the various analytical outputs, including images and visualizations, into a single user-friendly document in HTML. This is done to enhance the accessibility and usability of our findings for a diverse audience, including stakeholders and urban planners. A brief description of the software and relative packages are explained in the following subsections.

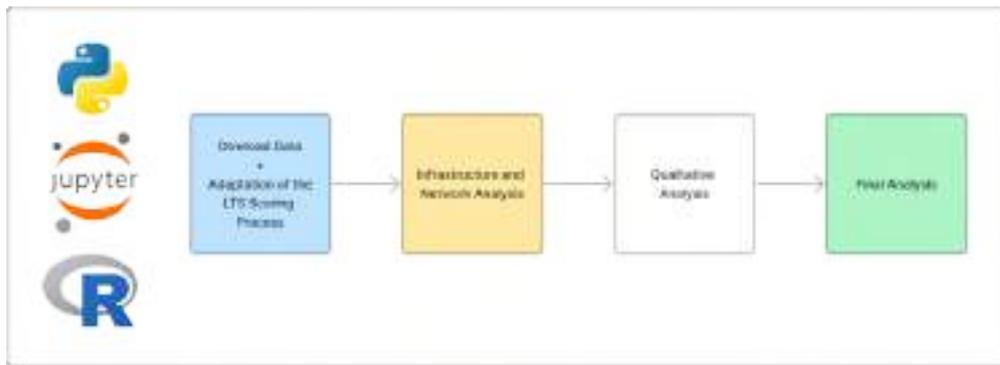


Figure 3.4: Diagram Illustrating the General Framework of the Analytical Tool

3.2.2 Software

All the scripts are run on three main environments, namely Python, R and Jupyter Notebooks.

Python is a high-level interpreted programming language known for its simplicity and readability, making it a popular choice for beginners as well as experienced developers. It was created by Guido van Rossum and first released in 1991 [234]. Python's design philosophy emphasizes code readability through the use of significant whitespace and a syntax that allows programmers to express concepts in fewer lines of code than might be used in other languages. Unlike compiled languages, it is executed line by line, which make the debugging and testing processes easier. Moreover, it is dynamically typed, which means that the type of the variables are determined at run time and it has a large standard library with modules and functions. It runs on a wide range of operating systems and is extended by multiple libraries that address different questions in various fields and topics. Thanks to its flexibility, Python is used in diverse areas such as web development, scientific and numerical computing, artificial intelligence, machine learning, data analysis and more. There are different types of Integrated Development Systems (IDEs) that are used for working with Python and executing the code, such as Jupyter Notebook, Visual Studio Code, Spyder and PyCharm etc. In this case, I used **VSCoDe** by Microsoft [161] as I find it more intuitive and familiar. Specifically my version is the 1.84.2. **Python** version 3.9.13 is used along with the following packages:

- **os (from Python Standard Library):** A module in Python's standard utility library, `os` provides functions to interact with the operating system. It includes functionalities like handling file and directory paths, creating and removing directories, and fetching the operating system's environment variables.
- **pandas (v. 1.4.4):** A powerful data analysis and manipulation library. It offers data structures like DataFrames and Series, which are ideal for handling and analyzing structured data efficiently.
- **numpy (v. 1.24.4):** Short for Numerical Python, `numpy` is fundamental for scientific computing. It is known for its array object, N-dimensional array, which is faster and more compact than Python standard lists. It also includes functions for performing complex mathematical operations.
- **scipy (v. 1.9.1):** An open-source library Python-based used for scientific and technical computing. `scipy` builds on `numpy` and provides modules for optimization, linear algebra, integration, interpolation and other special functions.

- **osmnx (v. 1.8.0)**: A library designed for downloading administrative boundary geometries and street networks from OpenStreetMap. It allows the creation, projection, visualization and analysis of street networks.
- **networkx (v. 3.2.1)**: A library for creating, manipulating, and studying the structure, dynamics, and functions of complex networks. It provides tools for the creation of graphs, network analysis and visualization.
- **geopandas (v. 0.14.1)**: An extension of pandas enabling spatial operations on geometric types. It is particularly suited for working with vector data because it simplifies the process of reading, analyzing and visualizing geographical information.
- **folium (v. 0.14.0)**: A library used to create interactive leaflet maps. It's built on the capabilities of the leaflet.js library and is especially useful in visualizing data that's been manipulated in Python on an interactive leaflet map.
- **matplotlib (v. 3.5.2)**: A plotting library. It provides a MATLAB-like interface and is useful for creating static, animated and interactive visualizations in Python.
- **seaborn (v. 0.11.2)**: A visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.
- **scikit-learn (v. 1.1.1)**: An open-source machine learning library for Python. It features various classification, regression and clustering algorithms, and is designed to interoperate with numpy and pandas.
- **imbalance-learn (v. 0.11.0)**: A Python package offering various re-sampling techniques commonly used in datasets showing strong between-class imbalance. It is compatible with scikit-learn and is part of the scikit-learn-contrib projects.
- **requests (v. 2.31.0)**: A Python HTTP library that makes it simple to send HTTP requests. It is used for making API calls, sending HTTP requests, and handling responses. It's known for its ease of use and ability to handle various aspects of HTTP.
- **shapely (v. 2.0.1)**: A package for manipulation and analysis of planar geometric objects. It is based on the widely deployed GEOS library and provides features like creating, manipulating, and analyzing complex geometric shapes.
- **h3 (v. 3.7.6)**: A library for working with Uber's H3 geospatial indexing system. It can be used to efficiently store data and perform computations on a hexagonal grid, useful for tasks like spatial analysis, data visualization and building geospatial applications.
- **rpy2 (v. 3.5.13)**: A Python interface to R. It allows Python and R to communicate with each other, enabling the user to run R code from Python. This is particularly useful for statistical analysis and visualization, leveraging R's capabilities within a Python environment.
- **pickle (from Python standard library)**: A module in Python used for serializing and de-serializing Python object structures, also known as marshalling or flattening. It converts Python objects into a byte stream suitable for storage or transmission over a network.
- **shap (v. 0.44.0)**: It stands for SHapley Additive exPlanation and is a Python library that provides a way to explain the output of machine learning models. It uses Shapley values from game theory to explain the contribution of each feature to the prediction.

R as delineated by the R Core Team [202], is a programming language intricately crafted for statistical computing and graphical representation. Its compatibility spans a broad spectrum of operating systems, further enhanced by its remarkable extensibility through an extensive array of packages. These packages, under continual development and refinement by the global community, are adept at addressing an extensive range of data analysis requisites pertinent to a multitude of scientific disciplines. For interfacing and operationalization within R, the *rpy2* package stands out as a pivotal tool. It furnishes a Python interface to R, facilitating bidirectional communication between these two programming languages. Predominantly, R-Studio is employed as the official Integrated Development Environment (IDE) [209]. Renowned for its efficiency and user-centric design, R-Studio significantly augments the processes of code development, data analysis, and visualization. In this research, the deployment of R version 4.3.0 was pivotal. This version was judiciously supplemented with a carefully curated ensemble of specialized packages. Each package was selected for its distinct functionality and instrumental role in addressing the analytical challenges of this study. Collectively, these packages substantially broaden R's fundamental capabilities, enabling intricate data manipulation, comprehensive analysis, and bespoke graphical representation, all meticulously tailored to meet the specific demands and nuances of this research endeavor.

- **dplyr (CRAN v. 1.1.2)**: A package in R for data manipulation. It provides a set of tools for efficiently modifying and summarizing data frames, making it easier to perform common data manipulation tasks. 'dplyr' is known for its user-friendly syntax and speed.
- **sf (CRAN v. 1.0-14)**: The complete name is Simple Features for R. This package provides support for simple features, a standardized way to encode spatial vector data. 'sf' makes it easier to work with geospatial data, supporting operations like reading, writing, manipulating and visualizing spatial data.
- **stplanr (CRAN v. 1.0.2)**: A package designed for sustainable transport planning with R. It provides a set of tools for analyzing spatial lines, networks and movement data and is particularly useful in transport planning and geographic analysis.
- **raster (CRAN v. 3.6-23)**: An R package for handling raster (grid) data. It allows for the creation, manipulation, analysis of raster data and is widely used in remote sensing and spatial analysis.
- **slopes (CRAN v. 1.0.1)**: This package calculates the slopes of streets or paths from line and point features. It's particularly useful in topographic and transportation analysis where understanding elevation changes is crucial.
- **geodist (CRAN v. 0.0.8)**: A package for fast, dependency-free geodesic distance calculations. It provides tools for calculating distances between geographical points using various methods.
- **geojsonsf (CRAN v. 2.0.3)**: An package that converts GeoJSON objects to simple features (sf) and vice versa. It is useful in the context of geospatial data workflows where GeoJSON is a common format for data exchange.
- **lwgeom (CRAN v. 0.2-13)**: This package is an R wrapper for the 'liblwgeom' library, enabling lightweight geometrical operations. It's used for advanced geospatial operations that are not covered by the 'sf' package.

Lastly, in the conducted project, the use of the ‘.ipynb’ file format, an acronym for Interactive Python Notebook, is pivotal. These files are integrally associated with **Jupyter Notebook**, a renowned open-source web application [135]. Jupyter Notebook enables users to craft and disseminate documents that encapsulate live code, equations, visualizations, and narrative text. Within this context, the term ‘notebook’ or ‘interactive notebook’ is not merely a metaphorical reference but a tangible representation of a computational environment that seamlessly integrates executable code, visual outputs, and explanatory text in an interactive milieu. The intrinsic capability of these notebooks to facilitate real-time code execution is a cornerstone feature, allowing users to witness immediate results, iterate, and refine their code in an explorative and dynamic fashion. Furthermore, these notebooks are adept at incorporating a diverse array of content, ranging from formatted text and both static and dynamic visualizations to complex mathematical equations, conveniently rendered through LaTeX. Jupyter Notebooks are particularly esteemed in the domain of data science, serving as a versatile tool for various applications such as data cleansing, transformation, numerical simulation, statistical modeling, and machine learning, among others. In this specific instance, version 7.0.6 of the notebook was employed. The decision to utilize multiple ‘.ipynb’ files was guided by the necessity to meticulously verify the outcomes of various functions and operations, a step by step process facilitated by the straightforward and user-friendly nature of these notebooks.

3.3 Data Analysis

This section elaborates on the comprehensive methodology and statistical techniques employed to rigorously examine and validate the posed research questions. It delineates the analytical approaches and statistical methods utilized to ensure a thorough investigation of the hypotheses. Furthermore, the section provides an in-depth exploration of two meticulously chosen case studies. These case studies are instrumental in contextualizing the research within real-world scenarios, offering a practical dimension to the theoretical analysis. Additionally, there is a detailed exposition of the datasets employed in these case studies, highlighting their relevance and alignment with the objectives of the research. This comprehensive presentation of the data analysis methodology, coupled with the case study narratives and dataset descriptions, forms the cornerstone of the research’s empirical framework.

3.3.1 Data Download

A script for extracting geospatial data from OpenStreetMap (OSM) was pivotal. This performs a range of functions:

1. **Data Retrieval:** Utilizes `osmnx` to fetch OpenStreetMap data for a specified city or region, focusing on both building data and street networks. This includes filtering major roads using a predefined list of road types (see Fig. 3.1).
2. **Geospatial Analysis:** Employs `geopandas` and `numpy` for spatial operations and calculations. This involves converting data to a projected coordinate reference system (CRS) for accurate distance calculations (EPSG:32632), and categorizing streets based on proximity to buildings within certain distance quintiles. In the specific, the script calculates the centroid coordinates of each building within the study area. Then, the `NearestNeighbors` function from `sklearn.neighbors` is applied to these centroid coordinates. This function performs a nearest neighbor analysis, determining the proximity of each building to its closest neighbors within the dataset. As mentioned, the categorization step involves dividing the

distances into quintiles, which are statistical values that divide a set of observations into five equal parts. The quintiles help in classifying streets into categories such as 'very close', 'close', 'moderate', 'far', and 'very far', based on their distance to the nearest buildings. By using this classification is possible to understand the context of each street segment (urban vs countryside). Streets in close proximity (no more than third quintile) to buildings might be indicative of a more urban or densely built-up area, while those further away might suggest a suburban or rural context.

3. **Slope Calculation:** Integrates a custom SlopeCalculator function to compute the slope of streets, leveraging raster data for detailed topographical analysis. Raster data for the area of interest must be uploaded in the *data* folder and then the file loads it automatically. The SlopeCalculator function incorporates the rpy2 library to interface Python with R, facilitating the use of specialized R packages. It executes R-based slope calculation on street network data, transforming the data into a format suitable for R processing, then applying the slope calculation, and finally re-importing the processed data back into Python as an uploaded geopandas GeoDataFrame with two new fields, namely '*slope*' and '*slope_class*'. The slope is calculated by adding elevation data per vertex of the linestring geometries, resulting in a 3D point geometry. The package then computes the average gradient for each line feature. The slope values are then multiplied by 100 to convert them into percentage terms, providing a more intuitive understanding of the steepness. The script classifies the calculated slope values into meaningful categories, such as '*flat*', '*mild*', '*medium*', '*hard*', '*extreme*', and '*impossible*'. This classification facilitates easier interpretation and analysis of the slope data. Moreover, summary statistics for the slope values are computed, providing insights into the distribution and range of slopes within the study area.
4. **Visualization:** Uses folium for mapping the spatial data, including slope information, to produce interactive maps that visually represent the terrain and infrastructure characteristics of the area under study.

The final dataset comprises the following fields, each representing a specific attribute of road segments:

1. **highway:** The category of the road or path in the network, e.g., motorway, residential road, cycle path.
2. **tunnel:** Indicates whether the road segment is a tunnel (Yes/No).
3. **reversed:** Denotes if the direction of travel for the segment is opposite to the originally mapped direction (True/False).
4. **osmid:** The unique OpenStreetMap ID assigned to the road segment.
5. **name:** The common or official name of the road, if available.
6. **length:** The length of the road segment in meters.
7. **oneway:** Specifies if the segment is a one-way road (Yes/No).
8. **bridge:** Indicates if the segment is a bridge (Yes/No).
9. **maxspeed:** The maximum speed limit on the segment, usually in km/h.
10. **junction:** Information about the type of junction, if the segment is part of one.

11. **ref**: Reference number or code commonly used for the road segment.
12. **lanes**: The number of lanes on the road segment.
13. **service**: Describes the service type of the road, applicable for service roads.
14. **access**: Information on the types of vehicles or traffic allowed on the segment.
15. **width**: The width of the road segment in meters.
16. **context**: Additional contextual information about the road segment, if available.
17. **group**: Categorizes the road segment into a group based on certain criteria (e.g., geographical area, road type).
18. **slope**: The slope or gradient of the road segment, often expressed as a percentage.
19. **slope_class**: A classification of the road segment based on its slope.
20. **geometry**: Geospatial data representing the shape and location of the road segment.

Each field provides valuable information for understanding the characteristics and functionalities of road segments in the network.

3.3.2 Level of Traffic Stress Analysis

The file in question is designed to modify the original Level of Traffic Stress (LTS) scoring system, as established by Mekuria, Furth, and Nixon [160], to better fit the diverse characteristics of Italian urban and non-urban city networks. This adaptation is essential due to the distinct geomorphological variations (such as changes in slope) and the unique aspects of road types, boundaries, and the varying degrees of data completeness or absence on OpenStreetMap in the Italian context. The goal is to create a more accurate and contextually relevant LTS assessment for Italian cities. The entire process is summarized in Fig. 3.5.

The methodology for calculating the Level of Traffic Stress (LTS) involves the loading of the updated network graph and a series of analytical steps that are based on a BikePathAnalysis function. The approach for this model is based on what has been done by the BikeOttawa association and other researchers that included the analysis for intersections [24, 32].

1. **Biking Permission Analysis**: The process begins by segregating road segments where biking is allowed from those where it is not. The function in the code assesses road segments within a GeoDataFrame to determine where biking is permitted. It evaluates several conditions based on existing columns such as *'bicycle'*, *'footway'*, *'access'*, and *'highway'*. These conditions identify roads where biking is explicitly disallowed or where the infrastructure implies a prohibition (like motorways or footways designated as sidewalks without bicycle permission). The script categorizes each road segment under various rules based on these conditions. Finally, it segregates the GeoDataFrame into two subsets: one where biking is allowed and one where it is not, and returns these subsets.
2. **Path Separation Assessment**: The script identifies separated and unseparated paths, assigning the lowest LTS value to separated paths. The function checks for specific conditions:

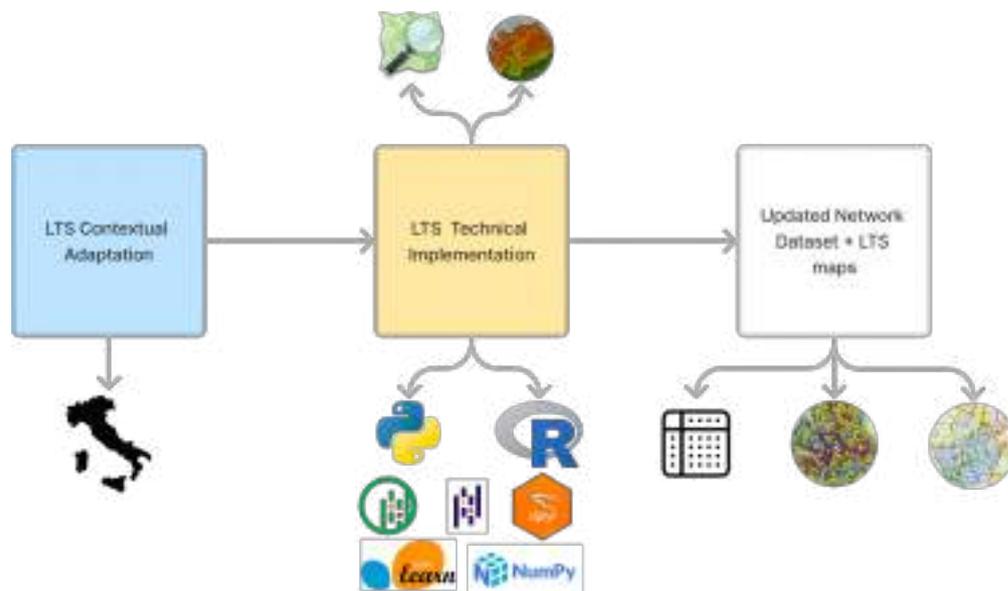


Figure 3.5: Diagram Illustrating the Level of Traffic Stress Adaptation Process

- **Cycleway Tags:** It looks for columns starting with *'cycleway'* to identify bike lanes. It sets conditions to true if a road segment is tagged as a *'track'* or *'opposite_track'*, indicating dedicated or opposite-direction bike lanes.
 - **Footway Condition:** It checks if a road segment is classified as a *'footway'* (a path primarily intended for pedestrians) but is not a crossing.
 - **Path Type Conditions:** The function then evaluates whether each segment matches any of the conditions set (being a cycleway, path, footway, having cycleway tracks, or opposite tracks) to classify them as separated paths.
3. **Bike Lane Analysis:** For unseparated paths, the presence of bike lanes is evaluated. The function in the code analyzes a GeoDataFrame to identify which edges (street segments) have bike lanes. It performs this by checking specific tags related to cycling infrastructure within the data. The function searches for columns that start with *'cycleway'* and checks if they contain identifiers like *'lane'*, *'left'*, *'opposite'*, etc., indicating the presence of a bike lane. Additionally, it considers *'shoulder:access:bicycle'* tags. The result is two subsets of the original data: one with edges having bike lanes and the other without.
 4. **Parking Impact Assessment:** The script checks for the presence of parking in bike lanes and adjusts LTS scores accordingly. The function works as follows:
 - **Identify Parking Tags:** The function first extracts all columns related to parking (e.g., columns with names starting with *'parking'*) from the given GeoDataFrame.
 - **Parking Detection:** It then determines which edges have parking based on specific identifiers like *'yes'*, *'parallel'*, *'perpendicular'*, *'diagonal'*, and *'marked'*.
 - **Dataframe Splitting:** The edges are split into two separate dataframes: one containing edges where parking is detected and the other containing edges where parking is not detected .

Then two functions perform an analysis of bike lanes considering the presence or not of parking and other roadway features. They follow a similar structure:

- **Lane and Speed Analysis:** They begin by determining the number of lanes and the maximum speed for each edge. To determine the maximum speed, it takes into consideration the different road types and relative limits in an Italian context, particularly when explicit speed limit data is not available. The typical speed limits are 90 km/h for national roads, 50 km/h for local urban roads, and 130 km/h for motorways. Then the function create a new column called *maxspeed_assumed*, and assigns the appropriate speed limit to each road segment.
 - **Width Analysis:** They check for the presence of a *'width'* or *'est_width'* column in the GeoDataFrame. If these columns are missing, default width values are assigned based on the type of highway and whether it is a one-way road. Specifically, they follow this scheme (*'motorway'*: 11.25m, *'primary'*: 7m, *'secondary'*: 6m or 5m in other cases. If is a *'motorway'* the width is divided in two).
 - **LTS Conditions and Assignment:** They define a set of conditions based on factors such as the number of lanes, maximum speed, and width of the road based on BikeOttawa's model rules [24]. Each condition is associated with a specific *'rule'* code (see Fig 3.2).
 - **Rule to LTS Mapping:** These *'rule'* codes are then mapped to corresponding Level of Traffic Stress (LTS) values (1,2,3 or 4). The conditions and corresponding LTS values are designed to reflect the varying degrees of stress a cyclist might experience under different road conditions (see Fig 3.2).
5. **Mixed Traffic Analysis:** An analysis of mixed traffic scenarios is performed, with particular attention to the influence of such conditions on LTS. The *mixed_traffic* function in the script evaluates various conditions to assign a Level of Traffic Stress (LTS) score to street segments. It begins by fetching data about lanes, width, and speed. Conditions are then set based on road attributes like highway type, motor_vehicle access, and maxspeed. Depending on these factors, different rules (such as *'m17'*, *'m13'*, *'m14'*, etc.) are applied to the street segments. Each rule corresponds to a specific LTS score, determined by a predefined rule dictionary (see Fig 3.2). The function assigns these LTS scores to the street segments, considering factors such as the presence of pedestrian paths, service roads and maximum speed limits. This approach provides a detailed and nuanced assessment of traffic stress for mixed traffic conditions.
 6. **Slope Penalty Application:** The script applies a slope penalty, adjusting LTS scores based on the steepness of the terrain. Specifically, the function in the code adjusts the Level of Traffic Stress (LTS) scores of road segments based on their slope and context (urban or non-urban). For each edge, it considers the slope category (flat, mild, medium, hard, extreme, impossible) and the segment's length. In urban areas, medium slopes increase the LTS by 1 if the segment is over 500 meters long, while hard and steeper slopes increase the LTS by 1 or 2, depending on length. In non-urban contexts, hard, extreme, and impossible slopes similarly increase LTS. The function ensures that the LTS score does not exceed 4. This adjustment reflects the additional challenge that slopes present to cyclists, especially in longer segments.
 7. **Node LTS Calculation:** The LTS score for intersections (nodes) is calculated, taking into account traffic control measures like stop signs and traffic signals. The function calculates the Level of Traffic Stress (LTS) for intersections (nodes) based on the highest LTS of adjoining road segments (edges). It first attempts to find the edges connected to a given

node, and if successful, it determines the maximum LTS from these edges. The LTS for a node is influenced by traffic control measures. If the node has traffic signals and the highest LTS from connecting edges is above 2, the LTS for the node is set to 2. Otherwise, if the node has a stop sign or traffic signals and the highest LTS is 1 or 2, the LTS for the node is set to 1. In all other cases, the node LTS is set to the maximum LTS of its intersecting edges.

This comprehensive process integrates various factors, including road type, traffic conditions, and topographical features, to generate a nuanced and context-sensitive LTS assessment.

Table 3.2: Description of LTS Rules

Rule Code	Description
p2	Cycling not permitted due to bicycle='no' tag.
p6	Cycling not permitted due to access='no' tag.
p3	Cycling not permitted due to highway='motorway' tag.
p4	Cycling not permitted due to highway='motorway_link' tag.
p7	Cycling not permitted due to highway='proposed' tag.
p5	Cycling not permitted. When footway='sidewalk' is present, there must be a bicycle='yes' when the highway is 'footway' or 'path'.
s3	This way is a separated path because highway='cycleway'.
s1	This way is a separated path because highway='path'.
s2	This way is a separated path because highway='footway' but it is not a crossing.
s7	This way is a separated path because cycleway* is defined as 'track'.
s8	This way is a separated path because cycleway* is defined as 'opposite_track'.
b1	LTS is 1 because there is parking present, the maxspeed is less than or equal to 40, highway='residential', and there are 2 lanes or less.
b2	Increasing LTS to 3 because there are 3 or more lanes and parking present.
b3	Increasing LTS to 3 because the bike lane width is less than 4.1m and parking present.
b4	Increasing LTS to 2 because the bike lane width is less than 4.25m and parking present.
b5	'Increasing LTS to 2 because the bike lane width is less than 4.5m, maxspeed is less than 40 on a residential street and parking present.
b6	Increasing LTS to 2 because the maxspeed is between 41-50 km/h and parking present.

Table 3.2 – continued from previous page

Rule Code	Description
b7	Increasing LTS to 3 because the maxspeed is between 51-54 km/h and parking present.
b8	Increasing LTS to 4 because the maxspeed is over 55 km/h and parking present.
b9	Increasing LTS to 3 because highway is not 'residential'.
c1	LTS is 1 because there is no parking, maxspeed is less than or equal to 50, highway='residential', and there are 2 lanes or less.
c3	Increasing LTS to 3 because there are 3 or more lanes and no parking.
c4	Increasing LTS to 2 because the bike lane width is less than 1.7 metres and no parking.
c5	Increasing LTS to 3 because the maxspeed is between 51-64 km/h and no parking.
c6	Increasing LTS to 4 because the maxspeed is over 65 km/h and no parking.
c7	Increasing LTS to 3 because highway with bike lane is not 'residential' and no parking.
m17	Setting LTS to 1 because motor_vehicle='no'.
m13	Setting LTS to 1 because highway='pedestrian'.
m14	Setting LTS to 2 because highway='footway' and footway='crossing'.
m2	Setting LTS to 2 because highway='service' and service='alley'.
m15	Setting LTS to 2 because highway='track'.
m3	Setting LTS to 2 because maxspeed is 50 km/h or less and service is 'parking_aisle'.
m4	Setting LTS to 2 because maxspeed is 50 km/h or less and service is 'driveway'.
m16	Setting LTS to 2 because maxspeed is less than 35 km/h and highway='service'.
m5	Setting LTS to 2 because maxspeed is up to 40 km/h, 3 or fewer lanes and highway='residential'.
m6	Setting LTS to 3 because maxspeed is up to 40 km/h and 3 or fewer lanes on non-residential highway.
m7	Setting LTS to 3 because maxspeed is up to 40 km/h and 4 or 5 lanes.
m8	Setting LTS to 4 because maxspeed is up to 40 km/h and the number of lanes is greater than 5.

Table 3.2 – continued from previous page

Rule Code	Description
m9	Setting LTS to 2 because maxspeed is up to 50 km/h and lanes are 2 or less and highway='residential'.
m10	Setting LTS to 3 because maxspeed is up to 50 km/h and lanes are 3 or less on non-residential highway.
m11	Setting LTS to 4 because the number of lanes is greater than 3.
m12	Setting LTS to 4 because maxspeed is greater than 50 km/h.

After the above steps, the geospatial data related to urban transportation networks are saved accordingly in GraphML format. The script specifically saves node data and a subset of edges data (related to traffic stress levels) in CSV format, then it converts a dataframe to a GeoDataFrame and sets its Coordinate Reference System (CRS) for accurate spatial data representation (EPSG: 4326). Then by using the 'osmnx' library it creates the graph from the GeoDataFrame, including details like osmid, lanes, name, maxspeed, LTS, slope, slope_class, rule, length, geometry, 'lanes_assumed', 'maxspeed_assumed', 'message', 'short_message'. A separate file then displays the two maps showing respectively the levels of traffic stress in the roads of the road network (roads and segments are coloured differently according to the level of traffic stress) and in a grid of 400 m hexagonal areas called H3 [104]. In this case the graph is called '*choropleth map*'. Specifically, in both cases a Folium map is created and centered on the mean latitude and longitude of the dataset. It integrates a geocoder for location searching and an established color-coded feature groups for different LTS. In the second cases, the script calculates the most prevalent LTS in each hexagon, indicating the level of cycling stress in that area.

In the next four sections, namely Exploratory Spatial Data Analysis, Network Analysis, Cluster Analysis and Gap Analysis, we will go on to perform a comprehensive infrastructural analysis of the road network. The whole process is summarized by the diagram in Fig. 3.6.

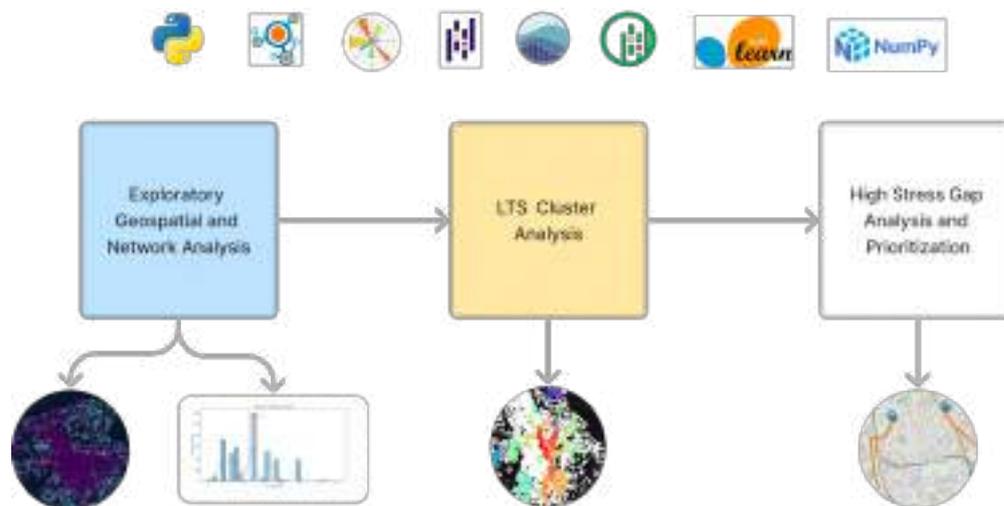


Figure 3.6: Diagram Illustrating the Infrastructure and Network Analysis Process

3.3.3 Exploratory Spatial Data Analysis

In this section, we explore the structural properties of the road network to gain a deeper understanding of the data. The analysis is performed using Python libraries such as OSMnx, NetworkX, and Matplotlib. The primary focus is on understanding the network's topology, its nodes, edges, and various graph properties. The road network data is loaded from a GraphML file, with its coordinate reference system projected to 'EPSG:4326'. This step ensures geographic consistency for spatial analysis. A directory structure is set up to store the output images and data derived from the analysis. The initial part of the analysis involves the extraction of basic graph properties such as the number of nodes and edges, giving a preliminary idea of the network's complexity. Then, a visual representation of the road network is generated. The roads are styled based on their type (*e.g.* motorway, primary, residential), with different colors and line widths. Several graph metrics are then calculated to provide deeper insights into the network structure:

- Degree of Nodes: The number of edges connected to each node is computed, indicating the connectivity of each node within the network.
- Network Density: This metric offers a view of how densely the nodes are interconnected.
- Triadic Closure: The concept of triadic closure is explored using the transitivity metric, which indicates the probability that two nodes with a common neighbor are connected.
- Other Statistics: Additional statistics like number of nodes, edges, the total length of all edges, the average node degree, network density, the average number of streets, the number of intersections intersecting at a node, the average circuitry and the proportion of edges that are self loops are calculated.

Then another function is designed to analyze and visualize the orientations of streets in the road network graph. This function provides insights into the directional tendencies of road layouts in a city or area by calculating the edge bearings. These are calculated as angles in degrees and they indicate the direction of the roads. Then it visualizes the frequency of edge bearings by using an histogram. The X-axis represents bearing degrees (0 to 360), and Y-axis shows the frequency of each bearing range. Finally, a polar plot is also created to visualize the distribution of street orientations in a circular format. The radial bars in the polar plot show the frequency of streets oriented within specific angular ranges. The plot is oriented with north at the top and rotates clockwise. This part of the analysis helps in understanding urban planning patterns, revealing predominant street directions.

3.3.4 Network Analysis

In this section, centrality measures are calculated for a road network graph to understand its structural properties. Centrality metrics are crucial in network analysis as they provide insights into the importance or influence of specific nodes and edges within the network [18, 30, 34, 65, 88, 174, 201, 217]

- Degree Centrality [30]: This measures the number of connections a node has. High degree centrality indicates a node with many direct connections, suggesting it as a significant junction or intersection in the road network. It identifies key intersections with the potential for high traffic flow or connectivity within the network.

- Closeness Centrality [34]: This metric calculates how close a node is to all other nodes in the graph, considering the shortest paths. High closeness centrality implies a node can quickly reach all other nodes, indicating central locations in terms of travel distance.
- Betweenness Centrality [88, 174]: It measures the extent to which a node lies on paths between other nodes. Nodes with high betweenness centrality often indicate critical points of passage or bridges between different parts of the network.
- Edge Betweenness Centrality [174] Similar to node betweenness, this metric is applied to edges. It identifies roads or segments that are frequently used in the shortest paths across the network, highlighting key connectors or bottlenecks.

Each centrality measure is calculated for the nodes and edges of the graph, then the top 10 nodes or edges for each centrality measure are identified, providing a focused view on the most significant elements of the network and visualized. Then, nodes and edges are visualized in a map by coloring them according to their centrality values. This graphical representation allows for an intuitive understanding of the network's structural properties. After this step, the Strongly Connected Components (SCCs) [178, 245] are calculated. This means identifying subgraphs where every node is reachable from every other node in the same subgraph, indicating areas of high interconnectivity. The results are also visualized to understand the spatial distribution of high-stress areas and the connectivity of the network. Specifically, nodes are positioned based on their geographical coordinates. Nodes and edges in the subgraph are drawn using red color to signify high stress. The style is set to distinguish these elements clearly against the background. Finally, the visualization is saved as an image file. With a second function, the top n largest SCCs in the graph are also visualized. The function first sorts the SCCs by size in descending order and save each one. Then centrality measures and node coordinates are compiled into a Pandas DataFrame, which is then saved as a CSV file. After this step, the road network graph is analyzed to assess traffic stress levels. The focus is on identifying and quantifying high-stress areas and understanding the distribution of traffic stress across the network. Understanding this is essential for urban planning and traffic management, as it highlights areas that might require infrastructural improvements or policy interventions. To identify them, nodes and edges with an LTS value of '4' are filtered and listed. The number of high-stress nodes and edges is then calculated and displayed, providing a quantitative view of the extent of high-stress areas within the network. After this step, LTS values for both nodes and edges are converted to integers for subsequent calculations. The average LTS for the entire network is determined together with the frequency of each LTS value for both nodes and edges. Two plots are then generated to visually represent the distribution of LTS among nodes and edges: the first plot shows the distribution of LTS values for nodes; the second one shows the distribution of LTS values for edges. Next, descriptive statistics, including mean, median, and standard deviation, are calculated for LTS values of both nodes and edges. These statistics provide insights into the central tendency and variability of traffic stress levels across the network. The same procedure is modified and repeated to consider the Level of Traffic Stress (LTS), for a road network graph. This method provides a nuanced understanding of the network's structure, emphasizing the impact of traffic stress on connectivity and accessibility. Specifically, all the centrality measures are weighted by considering the LTS. In the LTS - weighted degree centrality, the degree of each node is multiplied by its LTS value. Nodes without LTS data are excluded. This approach highlights nodes with high connectivity under high traffic stress, indicating critical points in the network under stressful conditions. Regarding the LTS-Weighted Closeness Centrality, for each edge, the 'length' attribute is divided by the LTS value, creating a new weight. Edges without LTS data or with an LTS value of zero are skipped. Regarding LTS-Weighted Betweenness Centrality

ones, edge weights are also here adjusted by dividing the 'length' by the LTS value. The top 10 nodes and edges for each centrality measure are also here identified and visualized, showcasing the most significant elements under traffic stress conditions. Then, additional visualizations include histograms of centrality values and bar plots for top nodes/edges, providing a detailed distribution and highlighting the most central elements.

The last part of the network analysis, focused on the concept of intersection density. A mentioned earlier, this concept is a critical measure that reflects the density and connectivity of a street network [20]. Empirical evidence in urban planning and transportation studies consistently demonstrates a positive correlation between intersection density and the prevalence of biking or walking as a mode of transportation [154, 246]. The initial step is to simplify the road network graph and this was achieved using the 'consolidate_intersections' function from the OSMnx library, designed to retain real intersections by consolidating nodes within a 5-meter radius. This process excluded dead ends and reconstructed the graph to ensure continuous connectivity. The simplification aimed to model actual intersections more accurately, addressing the issue of closely spaced nodes that may represent a single intersection in real life. Then, both the original and simplified graphs were converted to Geopandas DataFrames to facilitate spatial analysis. Then a custom function, was employed to visualize both the original and simplified graphs. The function plots each graph separately, highlighting the differences between the original and simplified networks. Next, to examine the spatial variance in intersection density, two additional visualization methods were employed:

- **Hexbin Plot:** This method involved creating a hexagonal binning plot of the nodes from the simplified graph. The hexbin plot provides a density map of intersections, offering a clear visual representation of intersection concentration across the network.
- **Kernel Density Estimation (KDE) Plot:** The KDE plot, overlaid on the street map, was used to estimate the density of intersections. This approach offers a continuous density surface, providing a nuanced understanding of intersection distribution relative to the street layout.

The visualization techniques, both hexbin and KDE plots, served to effectively communicate the results of the analysis, offering, under this assumption, intuitive insights into the distribution of intersections and so highlighting which areas are more bikeable/walkable.

3.3.5 Cluster Analysis

In this section, a comprehensive cluster analysis is implemented to identify and visualize high-stress zones within urban road network. This analysis utilized clustering algorithms, specifically DBSCAN [81], OPTICS [11], and HDBSCAN [157], to detect areas with a high level of traffic stress, indicative of potential challenges for urban mobility, particularly for cyclists and pedestrians. The graph model of the road network, is first stored in a GraphML file and loaded using the OSMnx library. Then, the graph is projected to the 'EPSG:4326' coordinate reference system for accurate geographical representation. After this, nodes with the highest Level of Traffic Stress (LTS 4), are extracted and insert in a GeoDataframe in order to enable spatial operations and visualizations. Then the clustering methods have been used:

- **DBSCAN:** In this method, we utilize the node coordinates as features. The determination of the *eps* parameter is guided by the k-distance graph method. More specifically, we plot a k-distance graph to ascertain the optimal *eps* value for DBSCAN. This graph displays, on the X-axis, points sorted according to the distance from their k-th nearest neighbor, and on the Y-axis, the distance to the second nearest neighbor. The process begins with

selecting a value for k , which signifies the count of nearest neighbors to evaluate for each point in the dataset. The identification of an appropriate radius (eps), within which points are deemed as neighbors, is facilitated by the elbow method. This method locates an elbow in the graph, typically indicating a pronounced shift in the distance to the k -th nearest neighbor. The Y-axis value at this elbow point is generally an effective choice for eps in DBSCAN. Points situated below this elbow are regarded as being within dense regions (clusters), whereas those above are likely to be noise or outliers. The minimum number of samples is set to 5, aiding in the identification of an optimal distance threshold for clustering high-stress nodes.

- **HDBSCAN:** for this method, suited for spatial data due to its ability to identify clusters of varying densities, shapes and handling noise, we also use as features spatial information. Regarding the algorithm setup, the minimum cluster size is set to 15 and minimum samples (sample density) to 5. These parameters dictate the minimum size of clusters and the sample density, respectively. The algorithm then assign cluster labels to each high-stress node.
- **OPTICS:** for this method, chosen for its ability to handle clusters of varying density and size, nodes data converted into a GeoDataFrame. X and Y coordinates are used as features because they represent the spatial locations of nodes. As parameters, the minimum samples and minimum cluster size are the same of the HDBSCAN. Together with the parameter ξ , which is a threshold that helps the algorithm determine what changes in data density should be considered significant enough to define a cluster they control the sensitivity of the algorithm to cluster formation. The value of ξ is 0.10 since we want to consider valleys in the reachability plot that exhibit at least a 10% drop in reachability distance as potential cluster boundaries. This threshold implies a sensitivity to smaller changes in density, potentially leading to the identification of more clusters, but it was a balance with the quality and specificity of the cluster themselves. Next, each node in was assigned a cluster label.

For all the algorithms, two graphs are created:

- **Scatter Plot:** The first stage of visualization involved a scatter plot of the clustered intersections, color-coded by cluster, providing an intuitive visual representation of high-stress areas.
- **Road Graph Overlay:** The second stage involved overlaying the clustered nodes onto the road network graph. This step contextualized the clusters within the broader network layout, enhancing the spatial understanding of high-stress zones.

Then, to measure and compare the results of the algorithm, the following validation metrics from Python's sklearn library are used.

- **Silhouette Score [208]:** This metric was computed to assess the consistency within clusters. A higher silhouette score indicates well-separated, cohesive clusters.
- **Davies-Bouldin Score [68]:** This score evaluates the separation between clusters. Lower values indicate better clustering.
- **Calinski-Harabasz Score [42]:** This metric evaluates the clusters' variance ratio, with higher scores indicating better-defined clusters.

3.3.6 High Stress Gap Analysis

Then, in this section the adapted to **LTS - IPDC (Identify, Prioritize, Decluster, Classify) procedure**, a novel approach to analyzing urban networks, particularly focusing on cycling infrastructure and its relationship with levels of traffic stress, is implemented. This method, based on the original IPDC framework [237], aims to systematically assess and improve cycling networks not by considering the presence of gaps between cycle tracks in a road network, but by considering the presence of gaps at a high level of traffic stress between sub-graphs at a low level of traffic stress in a road network. The entire procedure is summarized in Fig. 3.7.

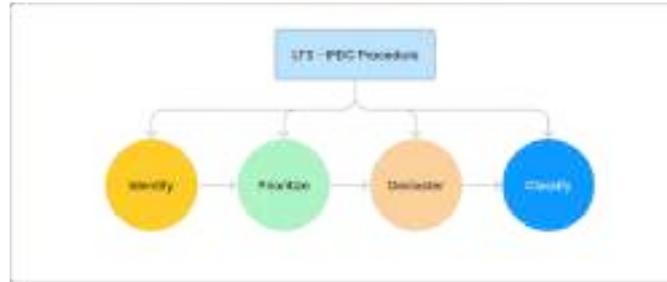


Figure 3.7: Diagram Illustrating the LTS - IPDC Procedure

The process begins with loading and projecting a graph representation of the road network using the *EPSG:4326* coordinate reference system. Then the graph is filtered to retain only low-stress edges and nodes, identified by 'lts' values of '1' or '2'. A new graph is then created from these elements. After this step, the graph is converted to an undirected graph to facilitate the identification of connected components, which are clusters of interconnected low-stress paths. The top 10 largest connected components are then identified and assigned unique colors for visualization purposes. Then, the entire network is visualized with different color codes to highlight the top 10 connected components. This step is crucial for a first visual analysis and identification gaps in the network. Subsequently, the method involves identifying gaps between the low-stress connected components. These gaps represent potential areas for improvement or expansion in the cycling network. Specifically, the procedure consists of 4 steps which are Identify, Prioritize, Decluster and Classify.

The **Identification** part is composed by the following steps:

1. **Identify Contact Nodes:** Contact nodes are defined as nodes where paths with different levels of traffic stress intersect or converge. Specifically, this step is about identifying where low-stress paths (lts values '1' or '2') come into contact with high-stress paths (lts values '3' or '4'). These nodes are critical for understanding transitions between different stress levels in the network. To identify them, we iterate through each node in the graph and for each node we examine its adjacent edges. These edges are the paths that emanate from or converge on the node. Then, the focus is on the 'lts' value associated with each edge to find the contact nodes. Nodes meeting the above criteria are added to the final contact nodes list, which will be used in the subsequent analysis of high-stress paths.
2. **Identify High-Stress Paths:** the next step involves mapping out high-stress paths between contact nodes using Dijkstra's algorithm [73]. This algorithm is selected for its efficiency in finding the shortest path in terms of distance, which is a critical factor for cyclists. A

path is considered high-stress if all of its constituent edges have *'lts'* values of '3' or '4'. We iterate through each edge in the path to check its *'lts'* value. If all edges meet the high-stress criteria, the path is added to the high stress paths dictionary created, with the source and target contact nodes as the key, and the path as the value. Finally, the output of this step is a collection of paths that represent the most challenging segments of the cycling network. This process is crucial for understanding the most stressful parts of the network and for prioritizing interventions.

3. **Filter Shortest High-Stress Paths:** This step focuses on filtering out the shortest high-stress paths from the previously identified high-stress paths in the cycling network. First, an empty dictionary, called *'shortest high stress paths'*, is created to hold the shortest path between each pair of contact nodes that have a high-stress connection. Then, we iterate through each high-stress path previously identified. For each path between a source and a target node, we check if the reverse path (from target to source) is already in the *'shortest high stress paths'* dictionary. This check is necessary because paths in a network can be bidirectional, and we want to ensure we are considering the shortest path in either direction. If the reverse path exists in the high stress path dictionary, we compare the lengths of the direct and reverse paths. The shorter of the two paths is retained in the *'shortest high stress paths'* dictionary. If there is no reverse path in *'shortest high stress paths'*, the current path is added directly. This step ensures that all unique high-stress paths are considered, even if they don't have a corresponding reverse path. As a result of this process, the *'shortest high stress paths'* dictionary contains the shortest high-stress path for each pair of contact nodes, considering both directions.
4. **Visualize Gaps:** the visualization step here involves two main components: visualizing edges with different levels of traffic stress and highlighting the gaps and contact nodes in the road network. First, each edge in the graph is assigned a color based on its *'lts'* value. Green represents low stress edges (*'lts'* values '1' or '2'), indicating safer, more comfortable routes for cyclists, while red represents high-stress edges (*'lts'* values '3' or '4'), indicating routes that are less safe or comfortable, typically due to higher traffic or other stress factors. Gray instead is used for edges without an *'lts'* value or with an undefined *'lts'* value, representing unknown or unclassified areas of the network. Then, the entire graph is plotted with edges colored accordingly. This visualization helps in quickly identifying areas of the network that are high or low stress, providing an at-a-glance understanding of the cycling infrastructure's quality. Then, the gaps, are highlighted in blue. This is done by iterating through the dictionary and changing the color of the edges in these paths to blue. To visualize the contact nodes, first the positions of all nodes are retrieved, then the nodes are drawn with a distinct blue color, differentiating them from the rest of the nodes. Finally, both visualizations are saved as images.
5. **Calculate Detour-Factor and Filter Gaps:** here, the *'Detour Factor'* to filter out certain gaps in the road network is calculated. As mentioned above, this is a metric used to evaluate the extent of the detour a cyclist must take to avoid a high-stress path and stay on a low-stress network. It is calculated as the ratio of the shortest path distance on the low-stress network to the shortest path distance on the entire street network. Here's a detailed breakdown of this process:
 - (a) **Iterating Through Shortest High-Stress Paths:** each pair of source and target nodes in the *'shortest high stress paths'* dictionary is visited. These pairs represents the start and end points of the high-stress paths previously identified.

- (b) **Calculating Shortest Path in Low-Stress Network:** for each pair of nodes, we calculate the shortest path distance on the low-stress network. This is the distance a cyclist would travel if they stick to low-stress routes between these two points. If no path exists in the low-stress network, the detour factor cannot be calculated and the gap is ignored.
- (c) **Calculating Shortest Path on the Entire Network:** the shortest path distance on the entire street network is also calculated. This distance represents the most direct route between the two points, regardless of stress levels.
- (d) **Detour Factor Calculation:** the detour factor for each gap is calculated. A higher detour factor indicates a longer detour on the low-stress network compared to the direct route, signifying a more significant gap in the network.
- (e) **Filter Gaps Based on Detour Factor:** gaps with a detour factor equal to or greater than a minimum threshold, in this case set to 2, are retained in the *filtered gaps* dictionary. This threshold is used to filter out gaps that don't significantly impact the overall route efficiency. Essentially, the focus is on gaps where the detour required to stay on low-stress paths is at least double the distance of the direct route.

Then, the next step is the **Prioritization** one. The goal here is to prioritize addressing gaps located on the most frequently used routes. As mentioned earlier, to effectively prioritize them, Vybornova et al. [237] suggest to consider the *link betweenness centrality weighted by gap length*. This technique operates on the premise that cyclists typically opt for the shortest path from start to finish. By applying the gap length as a weight to this centrality, the authors can estimate the total distance cycled in high stress gaps for each identified gap. In other words, the focus is on assessing and ranking the gaps in the road network based on their importance and potential benefit to cyclists. Here's a detailed explanation of this process:

1. Calculating Betweenness Centrality with Locality Parameter λ :

- (a) **Setting the Locality Parameter (λ):** the parameter is set to 2200 meters. This is crucial as it localizes the calculation of betweenness centrality to a specific range, focusing on gaps relevant within a certain neighborhood or area, rather than the entire city.
- (b) **Localized Betweenness Centrality Function:** this custom function computes betweenness centrality for each node, considering only the shortest paths that fall within the λ range.
- (c) **Normalization of Centrality Values:** Depending on the size of the area being analyzed (big city or town/village), the centrality values are normalized differently. Specifically, in the case of a big city or region, where the network size (number of nodes) is large, the normalization factor is calculated considering the total number of pairs of nodes in the network, excluding the node itself:

$$\text{Normalization Factor (Big City)} = \frac{1}{(N-1) \times (N-2)},$$

where N represents the total number nodes. This factor accounts for the fact that in large networks, the potential number of node pairs (and thus shortest paths) is significantly higher. The betweenness centrality value for each node is then multiplied by this normalization factor. In the second case, for smaller networks such as those found in towns or villages, the normalization factor is simplified due to the smaller size of the network:

Normalization Factor (Town/Village) = $\frac{1}{N-1}$.

These adjustments ensure that the centrality values are scaled appropriately for the network size, making them meaningful for comparative analysis and decision-making in both large and small urban settings.

- (d) Application to the Network: the localized betweenness centrality is computed for all nodes in the network, considering the weight of the paths.

2. Estimating Bicycle Traffic Flow and Computing Benefit Metric with Cost-Efficiency:

- (a) Calculating Total Length and Betweenness Centrality for Gaps: for each filtered gap, the total length and the sum of the betweenness centrality of the nodes within the gap are calculated.
- (b) Estimating Total Number of Expected Meters Cycled on the Gap: this is the product of the sum of the betweenness centrality and the total length of the gap. It estimates the total cycling traffic that would benefit from the gap closure.
- (c) Expected Meters Cycled Per Investment Unit (B(g)): this metric represents the efficiency of closing the gap in terms of the expected benefit per unit of investment. If the gap length is zero, the benefit is set to zero to avoid division by zero.
- (d) Ranking the Gaps: The gaps are ranked based on their benefit metric, B(g), in descending order. This ranking helps in identifying which gaps would provide the most benefit upon closure.
- (e) Visualize the Heterogeneity of Gap Closure Benefits: the final step involves creating a plot that visualizes the ranked gaps and their corresponding benefit metrics. This graphical representation highlights the variability in potential benefits across different gaps, aiding in decision-making regarding which gaps to prioritize for closure.

The third step is the **Declusterization** one. In the 'Declustering' phase of the IPDC procedure, the goal is to refine and prioritize the gaps in the cycling network by analyzing their relative importance and potential overlap. Here's a breakdown of the process:

1. Counting Initial Gaps: the number of gaps is counted before any processing. This provides a baseline for understanding the extent of the network's issues.
2. Setting a Cut-off Benefit Metric (B_g_min): a threshold value for the benefit metric is set (140). This value is used to filter gaps based on their calculated benefit, ensuring that only the most significant gaps are considered.
3. Filtering Gaps Based on Benefit Metric: gaps with a benefit metric below the threshold are excluded. This step reduces the number of gaps.
4. Creating a Network from Remaining Gaps: a new graph is created using the filtered gaps. This graph represent the network of significant gaps.
5. Decomposing into Disconnected Components: the network is then decomposed into disconnected components, representing clusters of gaps. This step helps in understanding how gaps distribution across the network.
6. Processing Each Component: each disconnected component is analyzed by computing all shortest paths within the component, calculating the benefit metric for each path. Then, paths are sorted by their benefit metric in descending order. The path with the highest

benefit metric that hasn't been removed yet is selected and added to the final list of gaps and its edges are removed from the subgraph to prevent overlapping with other paths.

7. Refining the List: The final list of gaps is refined to include only those that meet the benefit metric threshold. This step ensures that the focus remains on the most impactful gaps.
8. Visualizing Declustered Gaps: a map visualization is created to display the declustered gaps. This includes plotting different road types with distinct styles, highlighting the declustered gaps with unique colors and adding markers for gap locations.
9. Additional Processing: names of edges from the graph are extracted and matched with the paths in the final gap list by creating a support function. This aids in identifying specific streets or areas that correspond to the gaps.

The final step is the **Classification**. This phase in the IPDC procedure involves categorizing the identified gaps in the cycling network to better understand their characteristics and to aid in decision-making for infrastructure improvements. To do this, a function called `classify_gaps` is created. This takes a list of gap paths, a dictionary of edge names and the graph object as inputs. Then, it iterates over each gap and retrieves the unique names of the edges for the gap using a helper function defined at the previous step. This helps identify specific streets or bridges that constitute the gap. After this step, a dictionary is initialized for each gap with fields for intersections, bridge detection, type of gap and edge names. The function checks if any edge in the gap has a name containing 'ponte' (Italian name for bridge). If so, the gap is classified as a bridge. On the contrary, if the gap is not a bridge, intersection data is added. This includes the node ID and coordinates for each intersection in the gap.

The second part of the analysis is about finally visualize the classified gaps on a map. The gaps are color-coded based on their type (street or bridge) using a `gap_type_colors` dictionary. Each gap is plotted on the map with a line representing its path. The color of the line corresponds to the gap type. Finally, the map is saved as an HTML file.

This classification process adds a layer of contextual understanding to the gaps by identifying whether they are predominantly streets or bridges and by providing specific location details. The visualization on a map with color-coding and labels makes the data more accessible and interpretable.

In the next three sections, namely Urban Space Quality Assessment, BNA Score and Accidents Analysis we will go on to perform a qualitative analysis of the road network. The whole process is summarized by the diagram in Fig.3.8.

3.3.7 Urban Space Quality Assessment

As mentioned above, this step represents an initial effort to evaluate urban space elements that impact the cycling experience, such as the accessibility of Points of Interest and travel times by adapting the Novack, Wang, and Zipf [177]'s pedestrian-focused approach to a similar, bike-centric perspective. It is also known as routing analysis and is composed by the following steps:

1. Using OSMnx, we extract amenities in the network based on specified tags (cafes, shops, leisure facilities, etc.). The full list can be found in Table 3.3.
2. The network is set with a uniform biking speed (15 km/h), and travel times (15 min) are calculated for each edge.
3. The network's nodes and edges are converted to GeoDataFrames, and a Pandana network object is created for efficient network analysis.

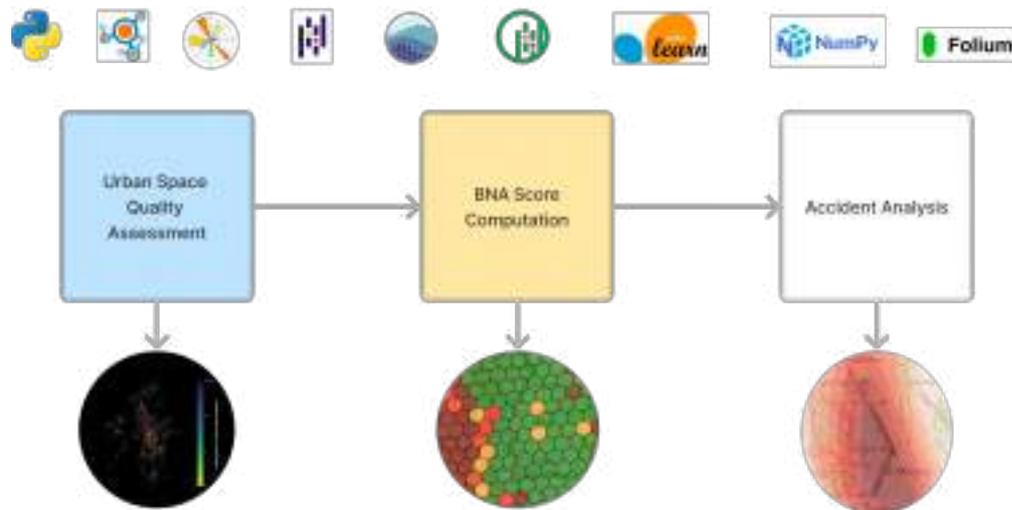


Figure 3.8: Diagram Illustrating the Qualitative Analysis Process

4. A function to visualize the biking time to the nearest POI for each node in the network is defined. The function creates a scatter plot where each point represents a node, colored according to the travel time. This provides a basic visualization of accessibility.
5. Another function utilizes Matplotlib's hexbins to create a more nuanced visualization. Hexbins aggregate nodes into hexagonal bins and display the average travel time for each bin, offering a clearer visual representation of accessibility across different areas.
6. The centroids of each POIs geometry are calculated and are set as POIs in the network, specifying the maximum travel distance for the analysis. Travel times to the nearest 10 POIs from each node are calculated. This provides the data for subsequent visualizations.

Methodological Significance:

- **Basic vs. Enhanced Visualizations:** the choice of using two plots has the following reason: an initial scatter plot provides a basic view of accessibility, but it can be crowded and less informative. The hexbin approach aggregates this data, making it easier to discern patterns and compare different areas.
- **POI Count Consideration:** recognizing that merely showing travel times to the closest POI might not reflect the true availability of amenities, we extend the analysis to include travel times to the 3rd and 6th nearest POIs. This approach provides a more differentiated understanding of accessibility.
- **Bikeability Assessment:** the final visualizations highlight areas that are easily accessible by bike and show how this *'bikeability'* varies depending on the method (nearest POI vs. 3rd/6th nearest POIs). This approach can reveal correlations with intersection density and other urban factors.

Category	Tags
amenity	cafe, bar, pub, restaurant, hotel
shop	bakery, convenience, supermarket, commercial, mall, department_store, clothes, fashion, shoes, retail, warehouse
leisure	fitness_centre, sports_centre

Table 3.3: Tags used for POI extraction

3.3.8 BNA Score

The Bike Network Analysis (BNA) score, as developed by PeopleForBikes [191] and Abad Crespo [2], is tailored and refined for use in Italian networks. The adaptation process includes integrating demographic data and information on common destinations, covering a range of sites including service, educational, and recreational locations. This score is based on several components:

1. Street Network Classification: based on the level of traffic stress, which determines the existence of low-stress routes conducive to comfortable cycling.
2. Selection of Destination Types and their Relative Importance: encompassing a broad spectrum of potential destinations, each assigned a rank based on its perceived importance to a typical commuter.
3. The Score Itself: quantifying the ease of access provided by the network to these destinations.

Specifically, as illustrated in Fig. (3.9), each of these primary components consists of various sub-elements. The classification of the street network is pivotal in identifying whether a low-stress pathway is available, enhancing the comfort of cyclists traveling between an origin and a selected array of destinations. Then, destinations are chosen from numerous possibilities and prioritized based on the value commuters might attribute to them [2, 189, 191]. Finally, the process assesses the accessibility afforded by the network and assigns a numerical score for each local area and a total one. The local score reflects the convenience and feasibility for bicycle commuters to reach their desired locations from a specific starting point.

First, since the first step is already done, datasets regarding edges and nodes and their LTS levels are loaded and converted to GeoDataFrames. Then, the coordinate reference system for both is set to *EPSG:32632* and then reprojected to *EPSG:4326* (longitude and latitude). Next, a function is designed to classify each row (representing a part of the bike network) into 'low' or 'high' stress based on the LTS value. Low stress is assigned to LTS values 1 or 2, and high stress to values 3 or 4. Subsequently, a graph is created using NetworkX with the nodes and edges of the above mentioned datasets. Isolated nodes in the graph are identified and removed to simplify the network. Two subgraphs are then created, one for low-stress edges and another for high-stress edges. Next, using Shapely library, a boundary for the network area is generated using geometries derived from the edges dataset. The *'simplify(tolerance=0.001)'* method is applied to the *'geometry'* column of the edges GeoDataFrame. This method simplifies the geometries by reducing the number of points, making computations more efficient. The *'tolerance'* parameter specifies how much simplification to apply (measured in the same units as the GeoDataFrame's coordinate system). Then, the *'unary_union'* function is used on the simplified geometries. This

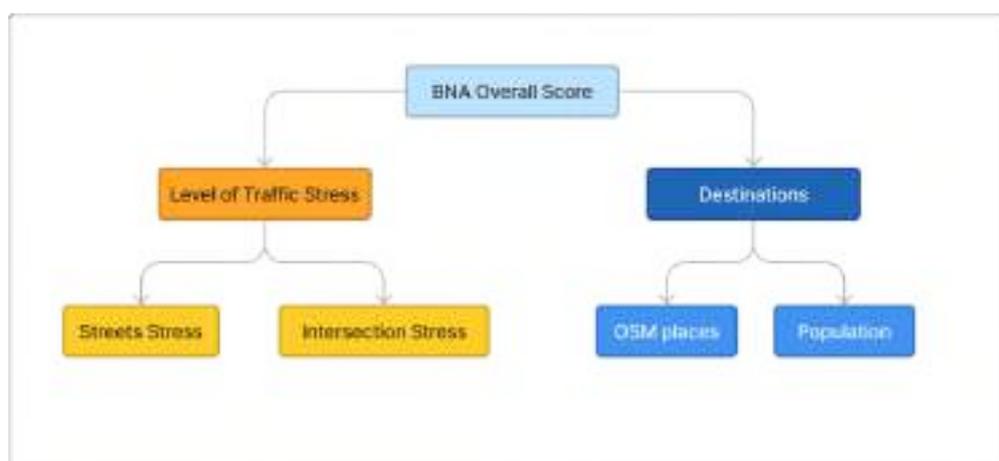


Figure 3.9: Diagram Illustrating the BNA Score Process

operation combines all the individual geometries in the dataset into a single geometric entity. It effectively merges overlapping or adjacent geometries and is useful for creating a unified spatial representation. A `.buffer(0.005)` method is then applied to the result of the unary union. This method creates a buffer area around the geometry. The parameter `'0.005'` defines the buffer's width (again, in the coordinate system's units). Buffers are typically used to create a zone around a geometry. Then, a function generates an hexagonal grid within the specified city boundary. Following the *H3: Uber's Hexagonal Hierarchical Spatial Index* [104] approach, it uses the `h3.polyfill` function to fill the city boundary area with hexagons of a specified resolution (9 in this case). Each hexagon is checked for intersection with the city boundary and only those intersecting are retained. The hexagons are stored in a GeoDataFrame with `EPSG:4326`. The hexagonal grid is spatially joined with a population GeoDataFrame to integrate population data into each hexagon. As mentioned above, the population GeoDataFrame is obtained from [136]. If an hexagon intersects with multiple population polygons, the populations are summed. The population data is then merged back into the hexagonal grid. Finally, a map is created and the hexagons are colored based on their population – red for high, yellow for medium, and green for low population. Red means that in these areas live more than the 2% of the population, while yellow more than 1% of the population. The next step involves the creation of a relationship between a hexagonal grid and the edges data of the network. The process aims to associate these edges with the hexagons they intersect. Specifically, a filtered version of the original edges GeoDataFrame is created containing geometry and other relevant attributes such as `u`, `v`, and `type_stress`. The `u` and `v` columns represent nodes at the start and end of each edge. Next, the grid GeoDataFrame is spatially joined with the new edges one using the `sjoin` function from GeoPandas. This join uses the `intersects` predicate, meaning an edge is associated with a hexagon if it intersects or lies within the hexagon. Next, the edges are grouped by each hexagon using the `groupby` function on the `hex_index` column. The resulting groups contain the `u` and `v` node pairs for each edge associated with a hexagon. A lambda function is then applied to transform these groups into lists of (u, v) tuples, representing the edges in each hexagon. Hexagons that only contain NaN values in their edge list are filtered out. Then, a subset of the new GeoDataframe is created where the `type_stress` is `low`. This subset is then grouped by hexagon, similarly to previous grouping. The groups of low stress edges are transformed into lists of (u, v) tuples, representing the low stress edges in each hexagon. Also here, hexagons with only NaNs are removed. After this

step, the shortest paths withing low-stress are computed. For each hexagons the closest nodes in the network are found. This is done by first finding the centroid of the hexagon and then querying a tree built from the coordinates of nodes in the original graph, for the nearest node. Then, to understand the continuity of low-stress routes, we checked if the low-stress subgraph is connected. If is not, we identify separate connected components. Then, we filter the representative nodes (those closest to the hexagons' centroids) to include only those present in the low-stress subgraph. A function is then created to calculate the shortest path lengths in the low-stress network from a given start node to all other representative nodes. It uses Dijkstra's algorithm for path finding [73], considering only nodes within the same connected component. The shortest paths computed are finally saved in a JSON file, providing a persistent and accessible record of the shortest paths in the low-stress network. A similar procedure is designed also to compute the shortest paths in general. Then, the procedure analyze connections between hexagons, focusing on whether these connections are considered 'low-stress'. The analysis is based on comparing the shortest path lengths in the entire network with those in a sub-network designated as low-stress. Specifically, a dictionary will store information about whether a connection (path) between pairs of hexagons is considered low-stress. A detour threshold is set to 1.5, which is used to determine whether a low-stress path is acceptably longer than the shortest path in the full network. For each hexagon in the shortest paths, a new dictionary is initialized for that hexagon and for each target hexagon the length of the path in the full network is retrieved. The length of the path in the low stress network is retrieved, if it exists. A connection between hexagons is considered true (low-stress) if the low-stress path length is not greater than the full network path length multiplied by the detour threshold. Otherwise, it is considered false (not low-stress). Then, two counters are initialized to count the the number of low-stress and non-low-stress connections, respectively. Several queries are then created in order to get different types of destinations from OpenStreetMap (OSM) data for a specified network using the *'osmnx'* library. Specifically, this process involves querying specific types of buildings, leisure facilities, shops, and potentially bus stops, then combining and filtering these destinations to create a unique list. For all destinations, name and coordinates (latitude and longitude) are extracted. Moreover, the second kind of query is taken from a list of destinations based on the work of Abad Crespo [2], which can be seen in Table 3.4 together with the Category, Weight and Scoring Process.

After loading the destination categories that have been queried from OSM, a spatial join of the resulting GeoDataframe is performed with the hexagons. Next, the goal is to determine and analyze the destinations accessible within each hexagon of a hexagonal grid. Specifically, the purpose is identifying the destinations that are reachable within a certain distance from the center of each hexagon. For each hexagon, we calculate the distance to each destination using Dijkstra algorithm [73]. Destinations within a threshold distance (e.g. 2000 meters in this case) are considered accessible. Finally, accessible destinations for each hexagon are stored in the dictionary. This dictionary includes the total number of accessible destinations and details about each destination (name, type, geometry, and index). Then a second iteration is computed similarly to obtain accessible destinations within a low-stress subgraph of the network. The process involves building a spatial index for the low-stress subgraph, iterating through each hexagon in a dataset, and computing distances to destinations using Dijkstra's algorithm. This code extends the previous accessibility analysis by focusing on low-stress paths, providing valuable insights into the accessibility of destinations via safer, less stressful routes. Then we calculate a score for each hexagon in a grid based on the types and number of destinations within it. The scoring is guided by a set of criteria defined in a JSON file also taken from Abad Crespo [2] which can be seen here (Fig. 3.5). The process involves parsing the JSON data for easier access. The *'destination_weights'* and *'type_weights'* hold weights for categories and types of destinations. Then, a function calculates scores for each destination type based on the number of destinations

Category	Type	Weight	Scoring Process
Opportunity	school	30	A
	library	15	B
	kindergarten	25	A
	university	30	A
Core services	doctors	20	D
	dentist	10	D
	hospital	30	C
	pharmacy	10	D
	supermarket	30	E
Retail	retail	33.3	D
	shop	33.3	D
	mall	33.3	D
Transit	train_station	50	G
	bus_station	50	G
Recreation	park	25	B
	fitness_centre	25	B
	sports_centre	15	B
	restaurant	15	B
	bar	10	C
	cafe	10	C

Table 3.4: Categories, Types, and Weights for Destination Scoring

and their order of importance as defined in the JSON file, and then aggregates these scores for each hexagon. The total_score for each hexagon is weighted by its population relative to the total population in the area. To better visualize the distribution of the number of hexagons for the rounded total score of each hexagon a bar plot was created. Each bar represents the number of hexagons for a specific rounded total score. The height of the bar corresponds to the count of hexagons with that score. Finally, an overall score is computed for the area covered by the hexagons. In summary, this process evaluates the accessibility and distribution of various amenities within a city or area. It does so by scoring hexagonal sections of a grid based on the types and quantities of destinations they contain, weighted by both the importance of these destinations and the population density of each hexagon. Finally, to visualize total scores for hexagonal areas in the network, each total score is extracted from each hexagon. A Pandas Dataframe is created with columns for hexagon index, geometry, and total score. Then it is converted to a GeoDataFrame to enable spatial operations and the CRS is set to *'EPSG:4326'*. A map is created by using a list of hexagons geometries and total scores. Each hexagon is colored

based on its total score and plotted on the map. This allows an intuitive visual understanding of how these scores are distributed across the city.

Process	Criteria Description	Points
A	First low stress destination	30
	Second low stress destination	20
	Third low stress destination	20
B	First low stress destination	40
	Second low stress destination	20
	Third low stress destination	10
C	First low stress destination	70
D	First low stress destination	40
	Second low stress destination	20
	Third low stress destination	10
E	First low stress destination	60
	Second low stress destination	20
F	First low stress destination	70
	Second low stress destination	20
G	First low stress destination	60

Table 3.5: Scoring Processes for Low Stress Destinations

3.3.9 Accidents Analysis

A qualitative analysis was conducted on accidents in the road network focusing over the past five years. This examination aimed to uncover spatial patterns and correlations between these accidents and varying levels of traffic stress. After loading datasets for streets and nodes and relative LTS values, they are converted to GeoDataframes to enable spatial operations. A GeoDataFrame of accidents is also loaded and unnecessary columns are dropped. The first step is done for visualizing accidents locations in the map and each one is marked with a red circle. Then other visualization plots are generate using kernel density estimate (heatmap) and heatmaps. Subsequently, A 10-meter buffer is added to each accident location in accidents GeoDataFrame to account for spatial imprecision. Then, a spatial join is performed between the buffered accident data and traffic stress data for streets, identifying accidents occurring near different road types. Similarly, accidents are spatially joined to the street network nodes. Then, dataset is filtered and fields like *lanes* and *maxspeed* are filled with assumed values if missing. A bar plot is created to visualize the frequency of accidents by road type, highlighting the most accident-prone areas. Additional visualizations, such as bar plots and scatter plots, are created to explore correlations between the number of lanes, speed limits, and accident frequencies. Then, a function is defined to extract numeric values from the *lanes* column, calculate their average, and handle missing

values. This function is applied to the *lanes* column to standardize and clean the data. Similarly, another function is defined and used to extract numeric values from the *maxspeed* column, using the maximum value as the representative speed. Histograms with Kernel Density Estimation (KDE) are plotted for both *lanes* and *maxspeed* to visually inspect their distributions. Then, the Shapiro-Wilk test is conducted for both columns to assess the normality of their distributions. Moreover, due to potential non-normal distributions, the Spearman correlation coefficient is calculated between *lanes* and *maxspeed* to assess their relationship. Specifically, the dataset is subdivided by road type, and Spearman correlations are calculated for each subgroup to understand the relationship within specific road categories. For road types where Spearman correlation is not applicable (NaN), the distributions of *lanes* and *maxspeed* are investigated to understand the diversity of data. A variance check is also performed for both fields within each road type category. By cleaning and standardizing data, visually inspecting distributions and employing statistical tests for normality and correlation, we gain insights into the factors that may influence cycling safety in urban areas. The use of Spearman correlation, in particular, allows for robust analysis regardless of data distribution characteristics. Next, the focus is on identifying accident frequency analysis by considering the LTS values. Accidents are filtered to exclude undefined LTS values and then grouped by LTS. For each LTS value, the percentage of total accidents is calculated to assess the relative risk associated with each traffic stress level. Two bar plots are created to visualize the number and percentage of accidents corresponding to each LTS value. The first is composed by the different four LTS values and the percentage of accidents for each. Then second has only two columns showing low stress and high stress percentages of accidents. This categorization simplifies the analysis by clustering LTS values into broader groups. Overall, these plots provide insights into which levels of traffic stress are more prone to accidents. Additional bar plots are created to show the number and percentage of accidents in each stress level category, providing a clearer understanding of where most accidents occur. A similar analysis is also done for accidents near intersections, categorized by LTS values and stress levels. Next, statistical tests are used to determine the significance of the differences in accident frequencies, specifically the Chi-Square test [188]. The outcomes of the Chi-Square tests are interpreted, and pairwise comparisons are conducted for significant findings, offering a deeper understanding of which categories differ significantly in terms of accident frequency. Finally, the data is converted to a GeoDataFrame for spatial operations and centroids are calculated for each accident cluster. A hexagonal grid over the study area is created to analyze LTS distribution spatially using the *H3: Uber's Hexagonal Hierarchical Spatial Index* [104] approach. For each hexagon, we identify the prevalent LTS of roads within its boundaries. The using Python Folium we create an interactive map, visualizing the hexagonal grid and assigning colors based on LTS classes. Accidents data is then added to the map for visual representation.

3.4 Final Analysis

In this section we perform a final analysis to synthesize insights from previous analyses and outcomes and determine if there are any correlations among the results found. The entire process is summarized in Fig. 3.10.

The **first analysis is about finding a potential spatial relationship between high level of traffic stress gaps and accident occurrences**, so as to give a more complete view of the areas that need urgent improvements in terms of safety. In fact, according to Tobler's laws, *'everything is related to everything else, but near things are more related than distant things'* [162]. So, this mean that each location possesses a unique characteristic derived from its spatial

context within the broader system. A need emerges from this consideration, which is the necessity of employing techniques in geospatial analysis that duly acknowledge and integrate the spatial structure of data. In alignment with this principle, spatial autocorrelation analyses are also run. The analysis begins by loading pre-identified high-stress areas (gaps) and the road network. Isolated nodes are then removed to focus on the connected components of the network, as isolated nodes are irrelevant in spatial autocorrelation analysis. Each path in the `filtered_gaps` dictionary is transformed into a `LineString` object. This step is critical for spatial analysis, as it converts discrete path data into a continuous spatial object, allowing for more nuanced and spatially aware analyses. Then, the `'sjoin_nearest'` function spatially associates accidents with the nearest high-stress lines. This step is crucial to establish a spatial relationship between accidents and high-stress areas, allowing for subsequent analysis of patterns or clusters. Then, the Queen contiguity method is chosen for the spatial weights matrix as it is a common approach in spatial analysis for polygonal data. It helps define spatial relationships based on shared boundaries or vertices, which is suitable for line-based features like streets. Since the graph is composed by different separated subgraphs, we decided to analyze these disconnected components separately to avoid the pitfall of misinterpreting spatial autocorrelation results that could be skewed by the lack of connectivity in the network. The analysis of disconnected components separately ensures that spatial autocorrelation is measured accurately within each component. Local Moran's I is a powerful statistic for identifying spatial clustering of high or low values and outliers. It helps detect whether high accident counts are spatially clustered around high-stress areas or dispersed. The use of Local Moran's I, a standard method in spatial statistics, is justified due to its ability to identify local clusters and spatial outliers, which is central to understanding the spatial dynamics of accidents in relation to high-stress bicycle network gaps. Results are collected and stored in results list, which includes the component number, segment index, P-value, Z-score and quadrant for each observation. Then, the results from the Moran's I analysis are merged with the high-stress GeoDataFrame. Only statistically significant results (p-value < 0.05) are considered to ensure the reliability of the findings. Finally, the significant local Moran's I results are merged back with the original dataset. A first map is created for visualizing the results. In this one, every high-stress gap is color-coded based on its quadrant classification, providing a clear visual representation of areas with different types of spatial autocorrelation. In a second image instead we only visualize the last filtered and classified gaps.

Then, the **second analysis regards the prediction of high-risk accident areas in the network**. First, geospatial data with BNA total score, network centrality measures GeoDataFrame, and accidents data are loaded. Then, a spatial join between the hexagonal scoring data and accident data is performed. This operation counts the number of accidents occurring within each hexagon. Next, the centrality measures (degree, betweenness, closeness) are aggregated for each hexagon to understand the network's structural importance. These measures are crucial as they are related with accident likelihood [53, 223]. Since spatial autocorrelation is observed in road accidents, the model's predictors also include the spatial lags of the same. Specifically, spatial autocorrelation occurs when the value of a variable observed at one location depends not just on the characteristics of that location, but also on the values of the same variable at nearby locations. So, the inclusion of spatial lags allow the model to account for the influence of neighboring areas and potentially improving model accuracy [146, 147]. Total scores for each hexagon are then selected together with the mean centrality measures as features for our predictive model. These features capture both the quality of cycling infrastructure and the traffic flow characteristics of each area. To categorize geographical areas into 'high-risk' and 'low-risk' zones based on the frequency of accidents, a data-driven threshold was established. This threshold is critical for transforming the continuous variable of accident counts into a binary classification, which is instrumental for the modeling part. The threshold is determined using a

percentile approach. This method involves defining a specific percentile that effectively separates the top percentage of areas with the highest accident counts (high-risk) from the rest (low-risk). For this study, the 85th percentile was chosen. This implies that areas with accident counts in the top 15% are considered high-risk, while the remaining 85% are categorized as low-risk. The percentile-based threshold is selected to ensure a data-driven, grounded approach, reflecting the actual distribution of accident counts in the specific dataset. This approach also provides adaptability to different datasets' distributions. Moreover, the data-driven threshold of 15% is not only a product of statistical calculation but is also chosen for its alignment with common risk assessment practices, such as defining risk matrices, where a smaller, more manageable subset is often prioritized for intervention [198, 207]. Additionally, this approach is advantageous for aligning with the contemporary requirements of diverse urban planning strategies focused on sustainable mobility and bicycle infrastructure plans, which emphasize the need to prioritize the most pressing interventions initially [55, 58, 226]. Then, the dataset is divided into training and testing sets. Features are standardized to have a mean of zero and a standard deviation of one, ensuring that our model is not biased by the scale of the variables. Class imbalance in our target variable (*risk.category*) is addressed using techniques such as SMOTE, BorderlineSMOTE, ADASYN and SMOTEENN [19, 50, 108, 114]. When one class significantly outnumbers another, it can lead to biased model predictions favoring the majority class. This bias can reduce the model's ability to accurately predict the minority class, which is often the class of interest (like high-risk areas in this case). Using multiple methods to address class imbalance allows to find the most effective approach. These techniques generate synthetic samples for the minority class, balancing the dataset. SMOTE generates synthetic samples for the minority class by interpolating between existing minority instances. ADASYN generates synthetic samples, but it focuses more on the samples that are difficult to learn (those near the decision boundary). BorderlineSMOTE is a variant of SMOTE that only synthesizes new samples along the borderline where the majority and minority classes overlap. Finally, SMOTEENN, which stands for SMOTE + Edited Nearest Neighbors, combines over-sampling with under-sampling. It applies SMOTE to over-sample the minority class and then cleans the data using an under-sampling technique (ENN) to remove any overlapping samples between classes [19, 50, 108, 114]. Then, weights of classes are adjusted in the cost function of the classifier. Higher weights are given to the minority class, making the classifier pay more attention to correctly predicting these instances. So this method adjusts weights inversely proportional to class frequencies in the input data. Logistic Regression, Random Forest, and SVM are chosen as candidate models for our problem. Choosing Logistic Regression, Random Forest, and Support Vector Machine (SVM) as candidate models for predicting high-risk accident areas is a strategic decision based on the strengths and characteristics of these algorithms. Here are the reasons why these models are suitable: Logistic Regression is chosen due to its easily interpretable results, its efficiency in computation terms and its good results in linearly separable datasets [37]. On the other hand, Random Forest Classifier can capture complex, non-linear relationships between features without the need for feature transformation. It's very robust in terms of resistance to overfitting, it provides insights in terms of feature importance and it performs well in several range of applications and fields [186]. Support Vector Machine (SVM) can be effective in cases where the number of dimensions is greater than the number of samples, which is beneficial for complex urban datasets. it is also flexible to different datasets thanks to the use of different kernel functions and it focuses on maximizing the margin between classes, which can be advantageous for clear classification in imbalanced datasets [62]. The above models are commonly used in traffic predictive analysis [76, 77, 152, 172]. They are contained in a dictionary where keys are model names and values are dictionaries of parameter grids. For Logistic Regression, we vary the inverse of regularization strength and the *'solver'*. For Random Forest Classifier, we adjust *'n_estimators'* (number of

trees in the forest) and *'max_features'* (number of features to consider when looking for the best split). For Support Vector Machine, we experiment with different *'C'* values and *'kernel'* types. Then, we initialize Logistic Regression, Random Forest, and SVM models with the calculated class weights. This helps mitigate the impact of class imbalance by giving more importance to the minority class. *'GridSearchCV'* for hyperparameter tuning, combined with Stratified K-Fold cross-validation. Stratified K-Fold ensures that each fold is a good representative of the whole, maintaining the same percentage of samples of each class as the complete set [170, 219]. Then, for each class imbalance method (*'SMOTE'*, *'ADASYN'*, *'BorderlineSMOTE'*, *'SMOTEENN'*) the training set is resampled. This step generates synthetic samples of the minority class to balance the dataset. Each model is trained on the resampled dataset, and the best hyperparameters are identified based on the ROC-AUC score. The model with the best hyperparameters is selected for each class of models. Each selected model is evaluated on the test set. The classification report, confusion matrix, Precision-Recall AUC, and ROC-AUC Score are computed. These metrics provide a comprehensive view of the model's performance, particularly in terms of its ability to distinguish between the two classes (high-risk and low-risk areas). Finally, the best-performing model (Random Forest Classifier) is retrained on the entire dataset, including the resampled data. We then use this model to predict the risk category for each hexagon. SHAP (SHapley Additive exPlanations) is used to interpret the Random Forest model. SHAP values explain the impact of each feature on the model's output, providing insights into what influences the risk prediction [148]. The predicted risk categories are merged back with the hexagonal GeoDataFrame. This step is crucial for visualizing the predictions results on a map. Specifically, the map is created to display hexagons colored according to their predicted risk category (low, high).

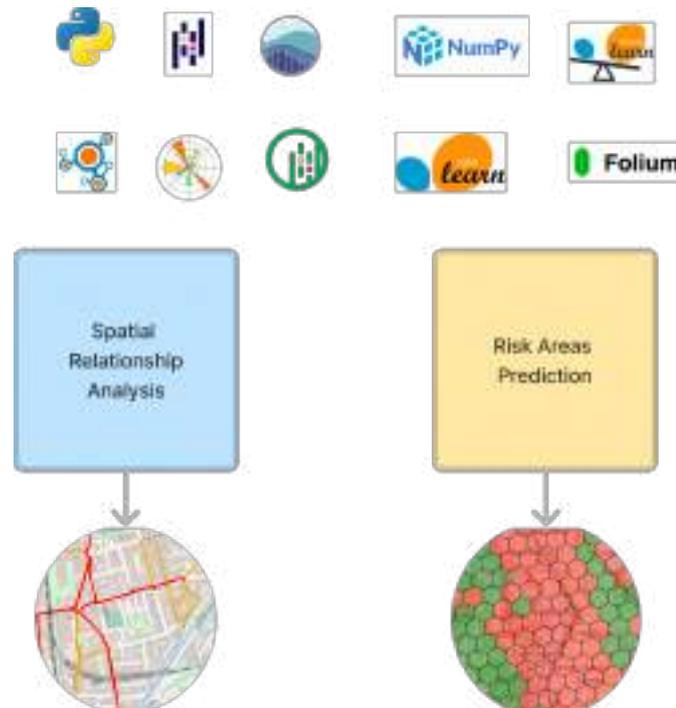


Figure 3.10: Diagram Illustrating the Final Analysis Process

Chapter 4

Results and Discussion

4.1 Results

In this section, we present and examine the outcomes of our analysis and evaluation conducted in the cities of Trento and Bolzano, our case studies. Initially, we focus on the adaptation outcomes of the Level of Traffic Stress (LTS) scoring process as applied to these two urban areas. Subsequent to this, we delve into the findings of the infrastructure network analysis, encompassing Exploratory Spatial Data Analysis (ESDA), Network Analysis, Cluster Analysis, and Gap Analysis. Following this, the third segment of our report details the results of our Qualitative Analysis. This includes an in-depth examination of Urban Space Quality Assessment, Bicycle Network Analysis (BNA) computation scores, and Accident Analysis. Finally, the fourth and last part tries to amalgamate finding from prior analysis to identify any potential correlations. Each of these findings is discussed within the framework of their respective limitations and contextual relevance.

4.1.1 Adaptation of Level of Traffic Stress Analysis

In examining the network of Trento through Figure 4.1, we observe a spectrum of traffic stress levels for cyclists, depicted through color-coded routes or hexagons. The coding ranges from LTS 1, signifying low-stress routes conducive to all cyclist demographics, to LTS 4, which corresponds to high-stress routes generally reserved for highly skilled cyclists. To elaborate:

- LTS 1 routes are displayed in green, highlighting low-stress corridors that accommodate cyclists across the spectrum of experience, from novices and children to seasoned riders.
- LTS 2 routes, rendered in light blue, denote areas of moderately low stress, ensuring comfort for a broad majority of adult cyclists.
- LTS 3 routes, illustrated in yellow, represent areas of moderate stress that could discourage less experienced adult cyclists.
- LTS 4 routes, marked in red, signify high-stress environments, typically navigated by cyclists with advanced skills.

The map analysis suggests that LTS 3 routes are the most prevalent, followed by LTS 2 and then LTS 4, with the latter being the least common. This observation is corroborated by Figure 4.2, which reflects a similar distribution of LTS levels across both intersections and

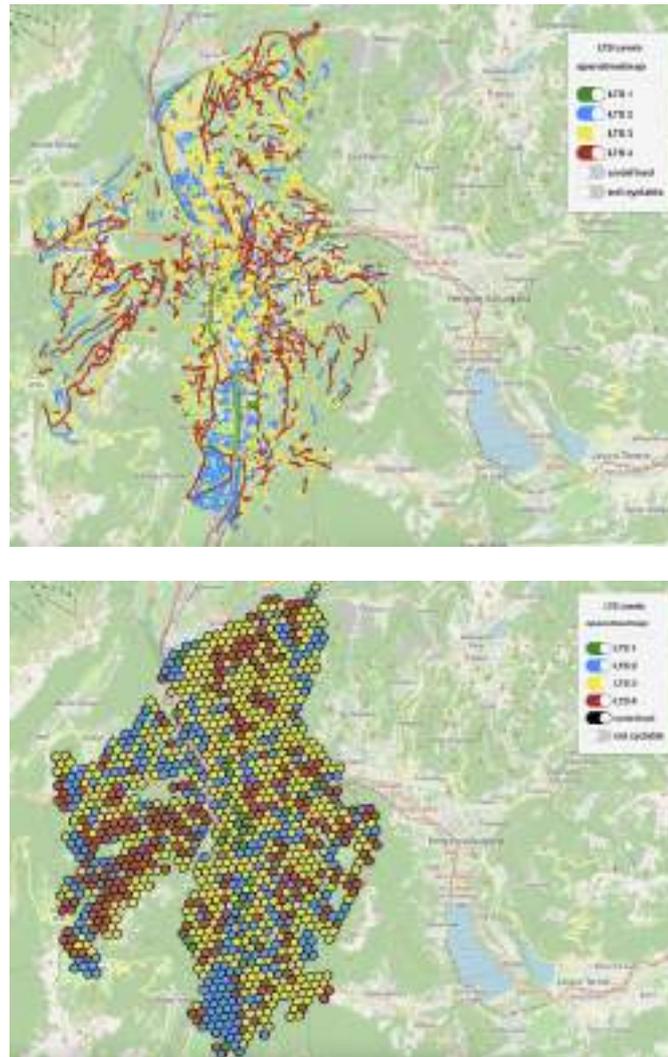


Figure 4.1: Top: Stress Network - Level of Traffic Stress in Trento. Bottom: Choropleth Map of Level of Traffic Stress in Trento.

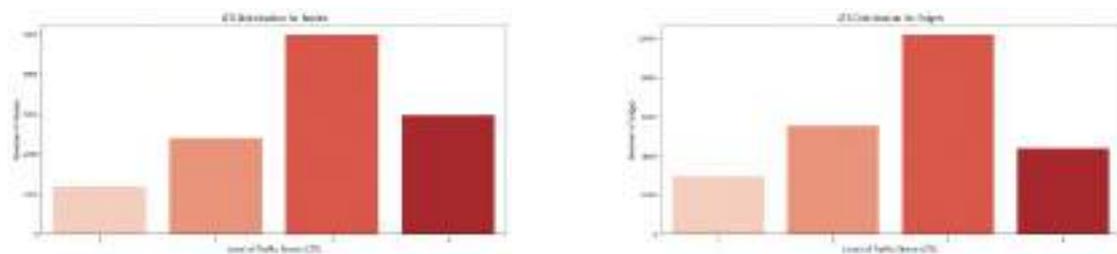


Figure 4.2: Left: Distribution of Traffic Stress in Trento's nodes. Right: Distribution of Traffic Stress in Trento's streets.

roadways. A notable distinction is identified in the roadway segments (edges), where LTS 2 is the second most common stress level, superseding the prevalence of LTS 4 which is more common in intersections. LTS 2 pathways are notably concentrated in the northern and southern peripheries of Trento, while their presence is more sporadic within the central urban area. Conversely, LTS 1 routes are prominently situated in the city center, an area characterized by restricted vehicular traffic, and are dispersed along a south-to-north gradient adjacent to established bicycle paths. Areas characterized by LTS 3 are dispersed throughout the city, likely indicative of busier streets lacking specialized cycling infrastructure, yet not reaching the severity of the highest stress level. In contrast, LTS 4 zones are most frequently found in the eastern and northern segments of the city. These regions, being elevated with winding, sloping roads and sparser routes, inherently possess a higher stress level and an increased risk perception, as deduced from the analysis which incorporates slope as a factor in the LTS assessment. The central zones and the one on the west side marked with LTS 4 correspond to areas with heavy traffic congestion (e.g Bassano Street, Laste Street, Venezia Square, Bernardo Clesio Street, 3 Novembre Corso). These insights facilitate a first understanding of the city's cycling dynamics, pinpointing specific areas where infrastructure enhancements are more needed.

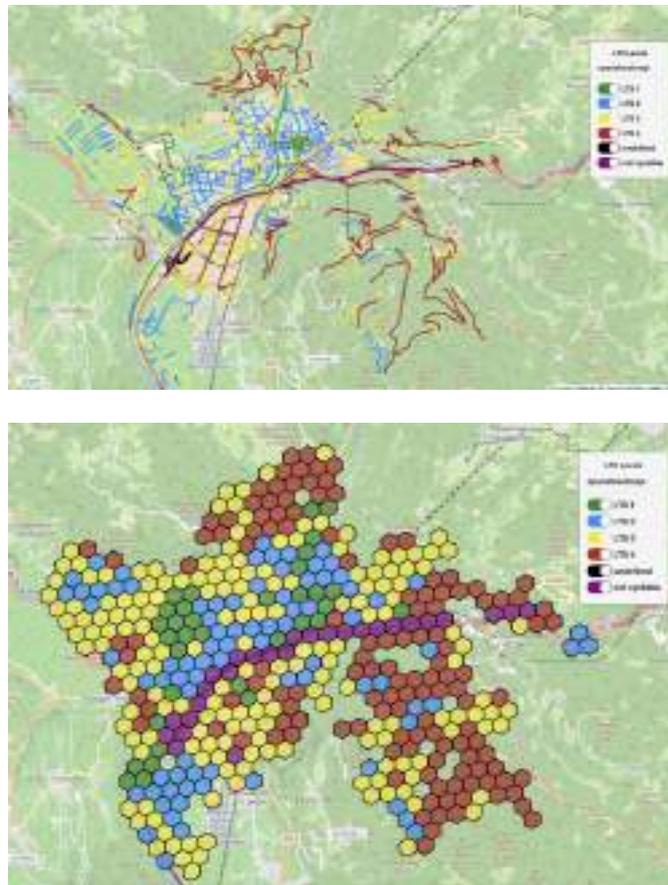


Figure 4.3: Top: Stress Network - Level of Traffic Stress in Bolzano. Bottom: Choropleth Map of Level of Traffic Stress in Bolzano.

Regarding Bolzano, the analysis of the map indicates that routes classified as LTS 3 are

predominant. These are closely followed by LTS 1 and LTS 2 routes. Notably, LTS 4 routes are primarily concentrated near mountainous areas, as it can be seen in Fig. 4.3. This observation is corroborated by Figure 4.4, which reflects a similar distribution of LTS levels across both intersections and roadways. A notable distinction is identified in the roadway segments (edges), where LTS 1 and LTS 2 are very similar.

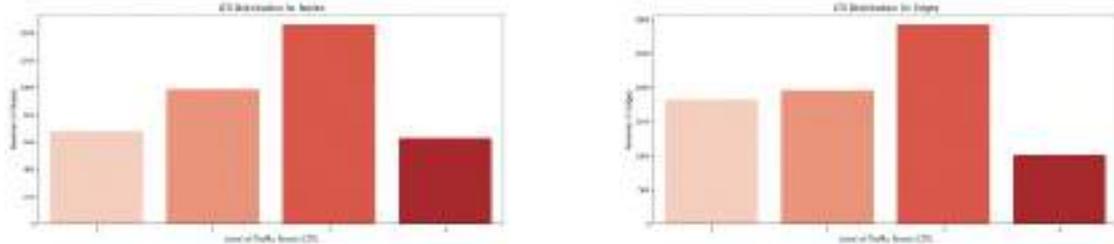


Figure 4.4: Left: Distribution of Traffic Stress in Bolzano's nodes. Right: Distribution of Traffic Stress in Bolzano's streets.

LTS 2 and LTS 1 pathways are notably concentrated in the center of Bolzano, but also in the western and southern areas. Interestingly, often LTS 1 routes often follow the paths of those at LTS 3 indicating the presence of bicycle route infrastructure near stressful routes, as it can be seen in Oltrisarco area which is mainly an industrial area. Areas characterized by LTS 3 are dispersed throughout the city but mainly in western and southern areas. LTS 4 areas are predominantly located in the eastern and northern parts of the city. Characterized by their elevated terrain and meandering, sloping roads, these regions naturally exhibit higher levels of stress and a heightened perception of risk, owing to their less dense route network and challenging topography. The other area with a lot of LTS 4 streets is the one above mentioned (Oltrisarco).

4.1.2 Infrastructure and Network Analysis

Exploratory Spatial Data Analysis

The exploratory data analysis of the Trento network yielded significant insights into its structure and characteristics. The network is represented in Fig. 4.5. The network consists of 21141 nodes and 23267 edges. It exhibits an average degree of 2.2 and a network density of $5e-05$, indicating a particular level of sparsity in the network's connections. Additionally, the triadic closure of the simplified version of the graph is calculated to be 0.032, suggesting a moderate level of interconnectedness among the nodes.

Further details of the network characteristics, such as the distribution of streets per node and other metrics, are encapsulated in the table above (see Table 4.1). Regarding Bolzano, its exploratory analysis yielded significant results too. The network is represented in Fig. 4.6. The network consists of 6134 nodes and 7736 edges. It exhibits an average degree of 2.52 and a network density of 0.00021, indicating a particular level of sparsity in the network's connections. Additionally, the triadic closure of the simplified version of the graph is calculated to be 0.032, suggesting a moderate level of interconnectedness among the nodes.

Further details of the network characteristics are encapsulated in the table above (see Table 4.2).

The orientation of our cities street network is also methodically examined to discern the directional preferences inherent to the city's roadways. This analysis not only elucidates the contemporary urban layout but also peels back layers of the city's historical growth narrative.



Figure 4.5: Trento - Street Network

The clear bias toward cardinal directions is reflective of modern urban design tenets, aimed at facilitating navigation and optimal land use. Conversely, any deviation from this regimented pattern may hint at the city's more historic segments or areas where the topography has dictated the road design.

Presented in Figure 4.7, the histogram of street bearings conveys the directional frequencies within Trento's network. The horizontal axis measures bearings from 0° , indicating due north, through a full 360° spectrum. Vertically, we see the occurrence of each directional bearing. Notably, the histogram displays distinct peaks, particularly prominent between 150° and 200° , and at the extreme ends of the polar spectrum, near 0° and 360° . These peaks betray a bimodal distribution, underscoring the preeminence of two primary orientations. This bi modality is emblematic of urban centers organized in grid-like configurations, with marked peaks proximate to the cardinal and ordinal points signaling a pronounced alignment with north-south and east-west axes.

Complementing the histogram, a polar plot in the same figure offers a radial perspective of Trento's street orientations. The circular display, akin to a compass, intuitively portrays the

Metric	Value
Number of Nodes	21141
Number of Edges	23267
Average Degree	2.2
Network Density	5e-05
Triadic Closure	0.03243
Total Edge Length	1869588.65 meters
Average Edge Length	80.35 meters
Average Streets per Node	2.76
Intersection Count	16655
Total Street Length	986480.15 meters
Street Segment Count	12800
Average Street Length	77.07 meters
Circuitry Average	1.1
Self Loop Proportion	0.0

Table 4.1: Summary of Exploratory Data Analysis Metrics for Trento



Figure 4.6: Bolzano - Street Network

bearings of the streets. Color segmentation within this plot is meticulously aligned with varying bearing ranges, enhancing interpretability. The radial lengths, particularly those extending towards the north and east, reveal the most frequently occurring street directions. This pattern is

Metric	Value
Number of Nodes	6134
Number of Edges	7736
Average Degree	2.52
Network Density	0.00021
Triadic Closure	0.03189
Total Edge Length	673731.5 meters
Average Edge Length	87.09 meters
Average Streets per Node	2.8
Intersection Count	4948
Total Street Length	376187.74 meters
Street Segment Count	4436
Average Street Length	84.8 meters
Circuitry Average	1.1
Self Loop Proportion	0.0

Table 4.2: Summary of Exploratory Data Analysis Metrics for Bolzano

indicative of deliberate planning, where road orientation serves the dual purpose of facilitating navigation and maximizing land utilization. The diversity in radial heights also intimates the existence of roads that meander away from the primary grid, painting a picture of a cityscape that has evolved organically alongside planned expansions.

In the context of Bolzano, the street bearing histogram elucidates the prevalent directions of the city's thoroughfares, as depicted in Fig. 4.8. Notably, there is a pronounced peak around the 0° bearing, indicating a substantial number of streets that orient in the meridional (North-South) direction. The histogram exhibits a non-uniform distribution characterized by multiple peaks, particularly in the 150° to 200° range, and corresponding valleys, indicating diverse street alignments. Such a pattern likely emerges from a confluence of factors, including the city's geographical topography and its historical patterns of expansion and architectural evolution.

The polar representation, on the other hand, offers a broader perspective on the variation in street orientations, highlighting a conspicuous bias towards certain bearings. A dominant peak is observed, suggesting a prevalent alignment of streets along the North-East to South-West axis, as evidenced by the elongated sectors pointing towards the 45° and 225° bearings. Additionally, the spread of street orientations from east to west is indicative of the city's adaptation to the valley's geographical contours. This demonstrates a cityscape that has been sculpted not only by human design but also by the natural landscape within which it resides.

The analytical value of these visualizations cannot be overstated; they are instrumental in deconstructing the urban fabric of Trento and Bolzano. Through the lens of these plots, we can observe the interplay between structured urban planning and the organic, sometimes serendipitous, development of the city's thoroughfares.

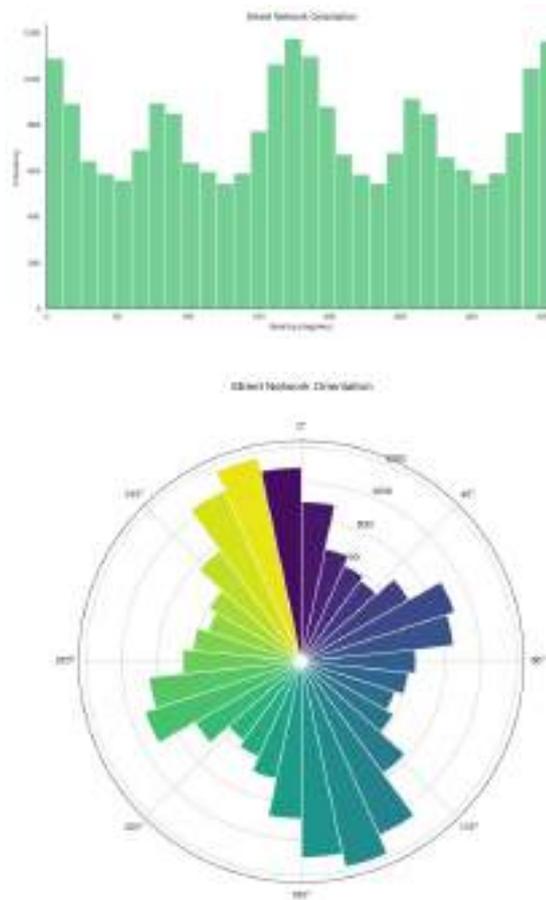


Figure 4.7: Top: Street Network Orientation in Trento. Bottom: Street Network Orientation - Polar Plot in Trento.

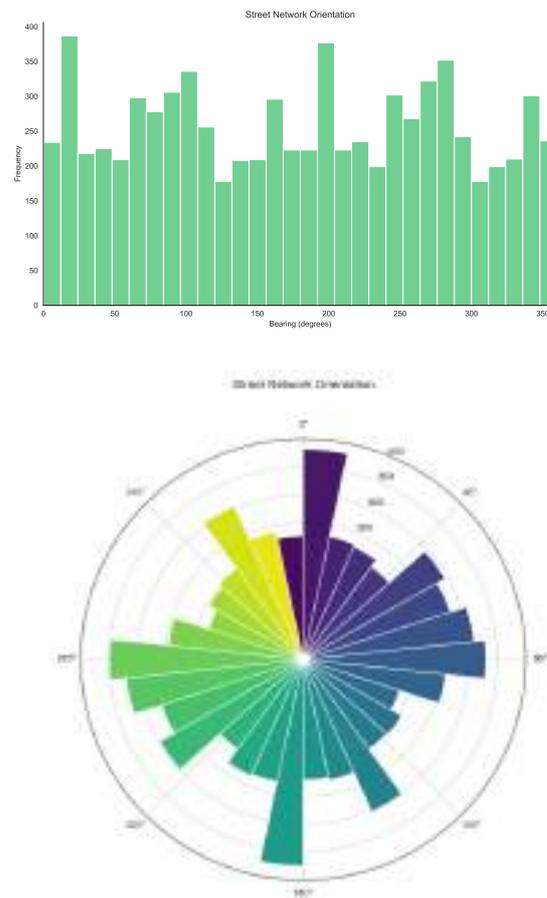


Figure 4.8: Top: Street Network Orientation in Bolzano. Bottom: Street Network Orientation - Polar Plot in Bolzano.

Network Analysis

Regarding centrality metrics, which are computed for the graph representation of a road network to decipher its structural attributes, these are the results for the two cities analysis, namely the degree centrality, the betweenness centrality, the closeness centrality and the edge betweenness centrality. These metrics are pivotal in the analysis of networks, offering a deeper understanding of the prominence or impact of particular nodes and links in the network's framework. Regarding the initial case study, namely Trento, the result, which can be seen here (Fig. 4.9), highlight that:

- **Degree Centrality:** on the top left is visible the graph representing the nodes of the city colored by their degree centrality. For enhanced clarity, nodes with no connections have been excluded. Each node is depicted as a circle of consistent size, with its centrality denoted by a chromatic gradient on the color scale to the right. This scale transitions from a deep purple, symbolizing the lowest centrality value of 1, to a vivid yellow, representing the highest centrality value of 9. The central area of the graph is densely populated with nodes, and the colors here are predominantly in the mid to high range of the scale, indicating a high degree of centrality. There are peripheral clusters of nodes with lower centrality, indicated by their cooler colors, suggesting they are less connected compared to the central nodes. Except for the central part, which is between 5 to 15, the overall structure of the graph seems to be somewhat organic, without clear boundaries or clusters, which might indicate a complex network with many interconnections rather than separated sub-groups.
- **Closeness Centrality:** on the top right is visible the graph representing the nodes of the city colored by their closeness centrality. It can be interpreted as the average length of the shortest path from the node to all other nodes. Isolated nodes are also removed here for simplicity. The color coding is quantified on a color scale similar to the one of the degree centrality. However, the notation '1e-5' at the top of the scale suggests that the values are very small. A higher closeness centrality is typically indicative of a node being more central in the network, as it would have, on average, shorter paths to all other nodes. This is visually represented by the bright yellow and green nodes located predominantly in the center of the graph, suggesting these nodes have shorter paths to other nodes. The distribution of nodes shows a gradient from the center outwards, with the most central nodes in yellow, surrounded by green, cyan, and then purple nodes as the centrality decreases. This pattern can imply that nodes at the periphery of the network are less integrated into the network's overall structure, having longer paths on average to other nodes. The central city cluster is the sole area exhibiting high closeness centrality.
- **Betweenness Centrality:** On the bottom left is visible the graph that illustrates the betweenness centrality of the nodes within the city of Trento. Betweenness centrality is a measure in network analysis that quantifies the number of times a node acts as a bridge along the shortest path between two other nodes. In other words, is an indicator of the importance in facilitating the connection. For simplicity the isolated nodes here are also removed and the size is reduced to highlight better the nodes with values bigger than zero. They are colored based on their betweenness centrality, which is reflected by the color scale shown on the right edge of the image. This scale transitions from dark purple at the lowest end to bright yellow at the highest end, with a noted maximum value of 0.030. The spread of colors across the nodes indicates the distribution of betweenness centrality values throughout the network. Nodes with a higher betweenness centrality are seen towards the central area of the graph, depicted in colors transitioning from blue to cyan, green, and then to yellow. The overall picture is clear, since the majority of nodes are purple,

suggesting that many nodes have a lower betweenness centrality, playing a less critical role as intermediaries in the network's connectivity. As mentioned above, a few nodes, particularly in what appears to be the network's center, show a higher betweenness centrality, highlighted by their bright yellow color. These nodes are likely significant in the context of the network's efficiency and resilience, as they may influence the speed and robustness of transportation.

- **Edge Betweenness Centrality:** On the bottom right is visualized the edge betweenness centrality within the city of Trento. Edge betweenness centrality is a measure in network analysis that reflects the frequency with which an edge occurs on the shortest paths between pairs of nodes within the network. The edges of the network are color-coded to represent their betweenness centrality, and the corresponding scale is displayed on the right side of the image. The range here goes from a minimum of 0.000 with a maximum at approximately 0.025. Edges with higher betweenness centrality values are more critical in the network's structure as they often form part of the shortest paths between various nodes. These are displayed in brighter colors, such as yellow and light green, and are predominantly located in what seems to be the central area of the graph. This is also in line with the graph representing the nodes betweenness centrality. Indeed, also here the majority of edges are depicted in dark purple, indicating lower betweenness centrality values. This suggests that these edges are less frequently used in the shortest paths and hence are less critical in the overall connectivity of the network. The layout of the edges shows a complex network with a dense center and sparser connections as one moves outward. This could imply that the center of the network has higher traffic and connectivity, potentially corresponding to the central area of the city of Trento.

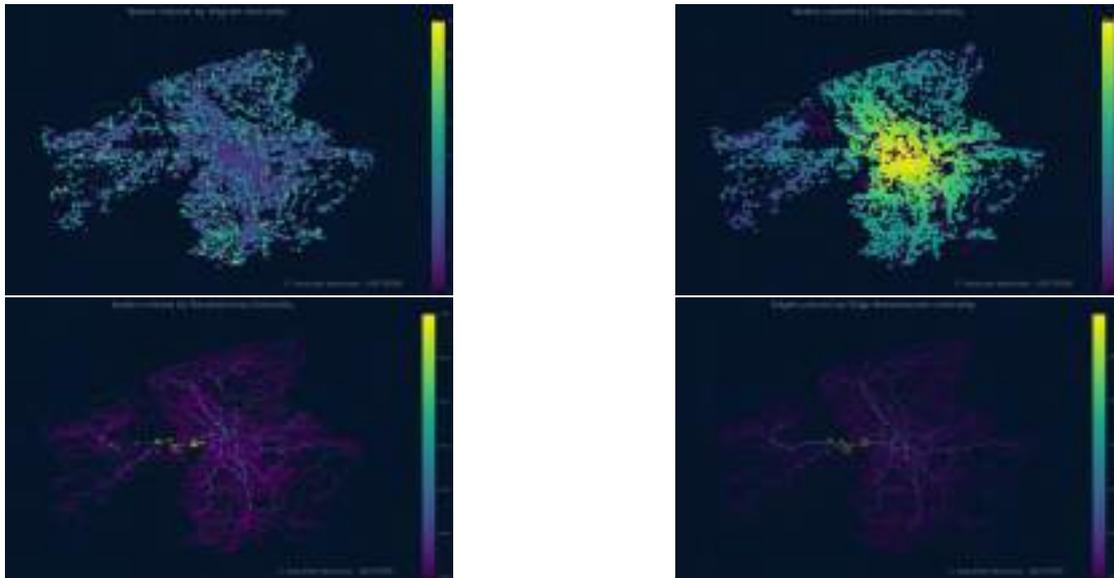


Figure 4.9: Top Left: Node Degree Centrality; Top Right: Node Closeness Centrality; Bottom Left: Node Betweenness Centrality; Bottom Right: Edge Betweenness Centrality — all within the Trento Network.

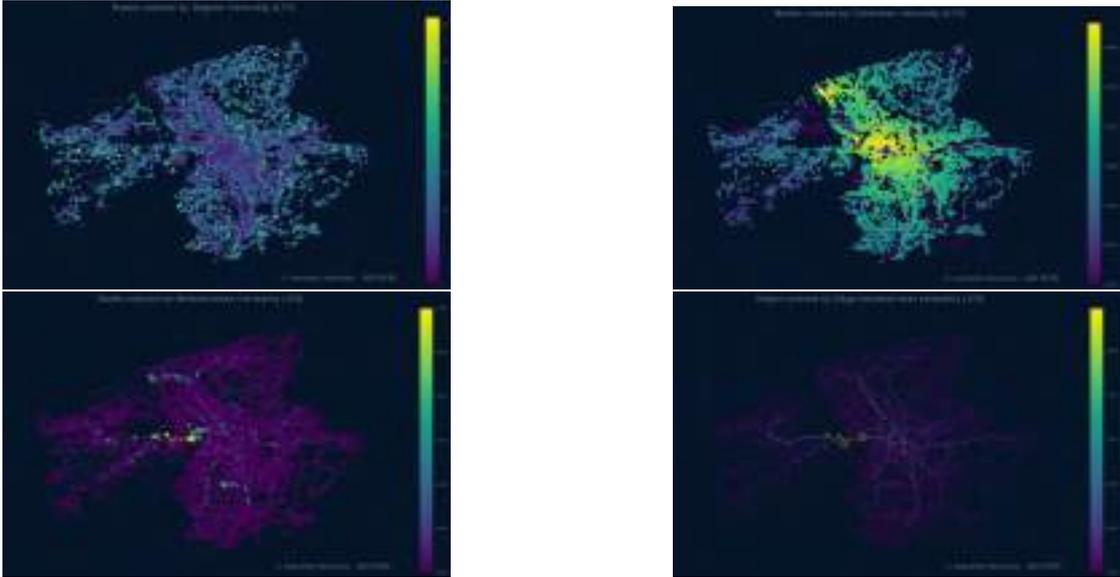


Figure 4.10: Top Left: LTS - Node Degree Centrality; Top Right: LTS - Node Closeness Centrality; Bottom Left: LTS - Node Betweenness Centrality; Bottom Right: LTS - Edge Betweenness Centrality — all within the Trento Network.

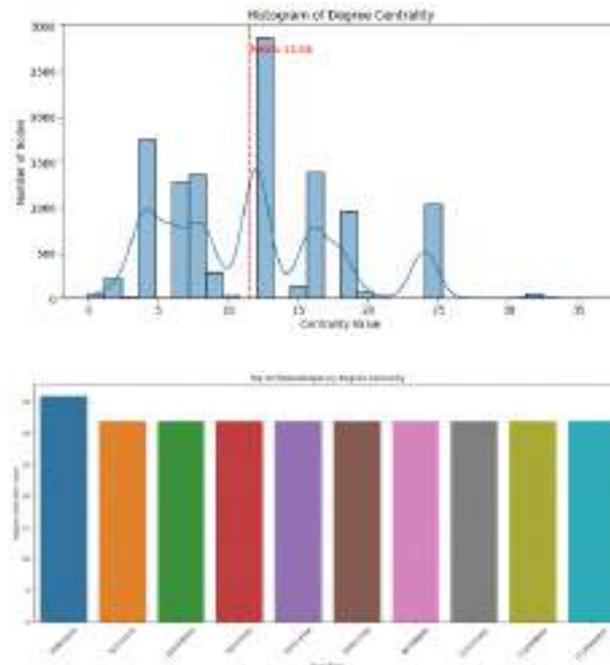


Figure 4.11: Top: Distribution of nodes for Degree Centrality values for Trento. Bottom: Top 10 nodes for Degree Centrality values in Trento.

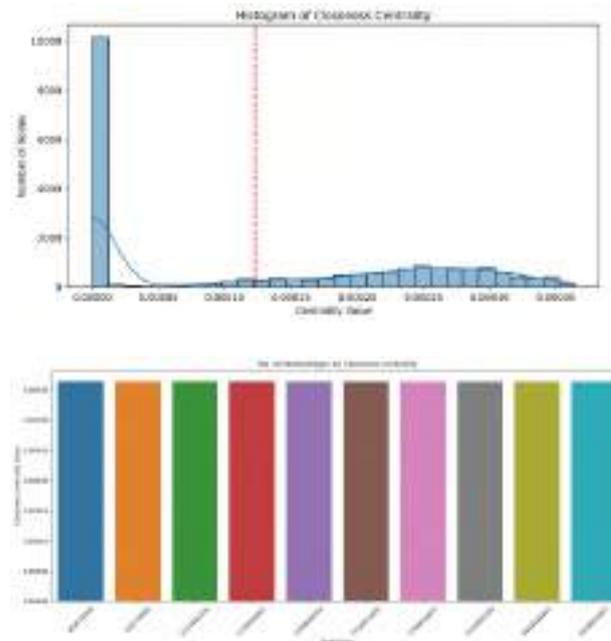


Figure 4.12: Top: Distribution of nodes for Closeness Centrality values for Trento. Bottom: Top 10 nodes for Closeness Centrality values in Trento.

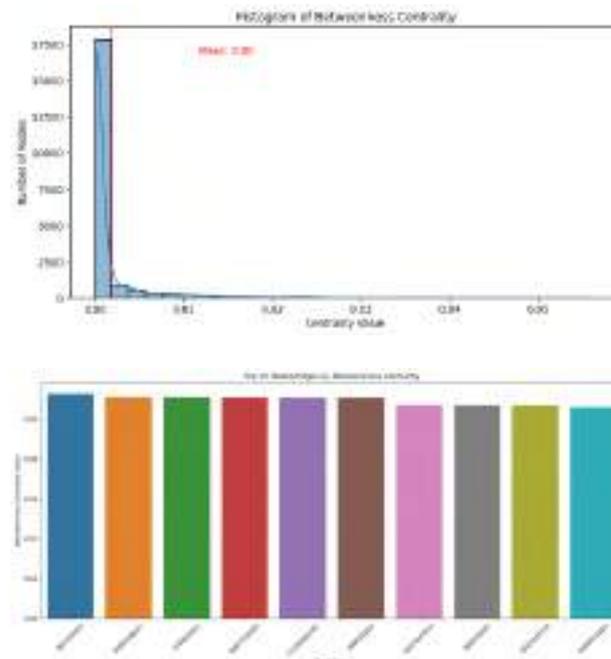


Figure 4.13: Top: Distribution of nodes for Betweenness Centrality values for Trento. Bottom: Top 10 nodes for Betweenness Centrality values in Trento.



Figure 4.14: High Stress Subgraph in Trento

With the availability of data pertaining to traffic stress levels, the same calculation was repeated using these as weights in the computation of centrality metrics. However, as can be observed from the graph (see Fig. 4.10), the outcomes remain visually extremely similar to those of the previous calculations.

The following outlines the descriptive statistical analysis of the centrality metrics distribution:

- Degree Centrality: on the top of Fig. 4.11, is depicted a histogram overlaid with a kernel density estimation (KDE) curve that represents the distribution of degree centrality values in the network. The horizontal axis represents different possible values of degree centrality within the network. The values range from 0 to over 35. The vertical axis indicates how many nodes possess a particular centrality value within the network. The bars represent

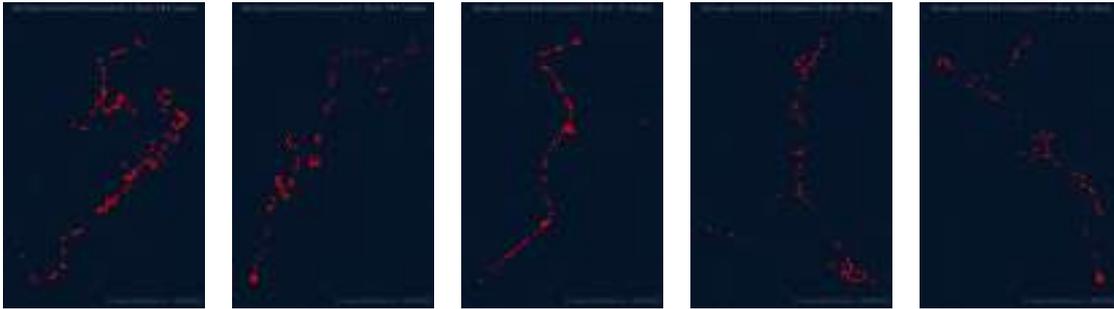


Figure 4.15: Top 5 High-Stress Connected Components in Trento

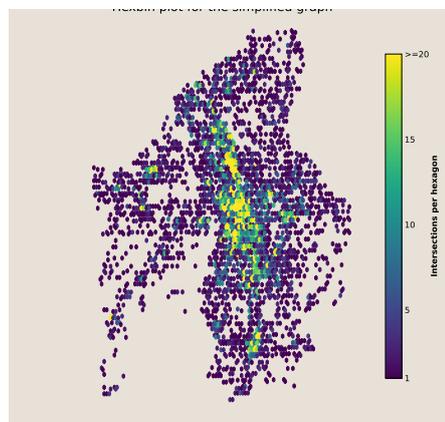


Figure 4.16: Hexbin map of Trento - Intersection Density

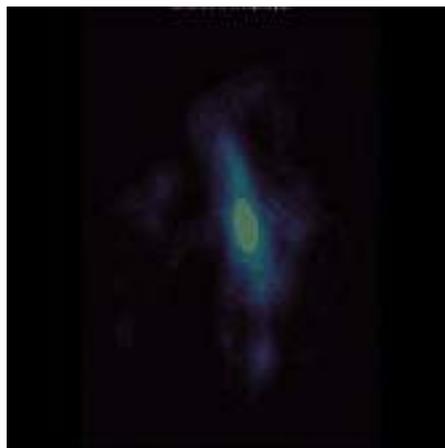


Figure 4.17: Kernel Density Estimation Map of Trento - Intersection Density

the actual distribution of the centrality values among the nodes, with the height of each bar indicating the number of nodes that have a specific degree centrality value. The tallest bar is near the centrality value of 10, suggesting that this is the most common degree centrality value among the nodes. The blue KDE curve, which is overlaid on the histogram, estimates the probability density function of the centrality distribution. It shows a smooth curve that highlights the underlying distribution pattern of the degree centrality values. A dashed red vertical line indicates the mean of the distribution. On average, a node in this network has a degree centrality value of approximately 11.56. The distribution appears to be right-skewed, meaning that a majority of the nodes have a lower degree centrality value, with fewer nodes having a high degree centrality. There are a few bars at the higher end of the centrality values, indicating that there are some nodes with significantly higher degree centrality. These could be important hubs or well-connected nodes within the network. On the bottom of the figure we can see the top 10 nodes for degree centrality values. The highest values is around 35 nodes.

- **Closeness Centrality:** on the top of Fig. 4.12, is depicted a similar histogram overlaid with the kernel density estimation curve. In this case it represents the distribution of the closeness centrality values. The values along the horizontal axis are very small, ranging from 0 to approximately 0.00035, which is typical when closeness centrality is normalized. There is a large spike in frequency at the lowest centrality value range, showing that a significant number of nodes have a closeness centrality close to 0. The dashed red line indicates a mean of the distribution, which does not coincide with the mode (the tallest bar) of the histogram, indicating that the distribution is not symmetric. The bars become progressively shorter as the centrality value increases in the first part, which suggests that fewer nodes have higher closeness centrality. Then in the second part there's an increase around 0.00025. The kernel density estimate provides a smooth curve that mirrors the general shape of the histogram but does not capture the initial spike. The distribution suggests that in this network, many nodes are not close to others on average, which might indicate a sparsely connected network or the presence of outliers or peripheral nodes with few connections.
- **Betweenness Centrality:** on the top of Fig. 4.13, is depicted a histogram that illustrates the distribution of betweenness centrality values. The distribution is heavily skewed to the left, with a significant number of nodes having a betweenness centrality value of 0, as indicated by the high bar at the beginning of the histogram. The mean of the distribution is 0.00, as marked by the red dashed line, which aligns with the first bar, affirming the left-skewness of the distribution. The frequency of nodes decreases sharply as the betweenness centrality value increases, showing that there are very few nodes with high betweenness centrality. Moreover, the histogram bars become increasingly sparse and lower in height as the centrality value grows, suggesting that nodes rarely serve as central bridges in the shortest paths between other nodes. The horizontal axis values ranges from 0 to 0.05, indicating the spread of betweenness centrality within the network and the y-axis indicates also in this case the count of nodes for each centrality value. The distribution indicates that the network may have a few key nodes that significantly influence the flow within the network, but the vast majority of nodes are not critical to such paths. This type of distribution is characteristic of networks where there are a few hubs or nodes that are highly central in terms of controlling the flow of information or resources, with most other nodes being relatively peripheral.

Subsequent analytical efforts were directed at identifying and quantifying areas subjected to high stress and understanding the distribution of traffic stress within the network. The observed

data, as illustrated in the Fig. (4.14), reveals a substantial subgraph of high-stress nodes within Trento. The majority of the nodes seem to be concentrated on the eastern part of the city, while they are more scattered on the western side. This observation can also be attributed to the fact that the network is less dense on the western side of the municipality, where there is a higher concentration of mountains. Overall, the density of high-stress nodes throughout the municipality indicates that the city experiences significant levels of traffic stress, a factor that should not be underestimated. Then, the first 5 high stress components are extracted (see Fig. 4.15). This is useful for visualize the core structures within the network, where the most interconnected nodes reside. As can be seen from the 5 top components, all of them intersect in the central area of the network, indicating that this hub must necessarily be considered during the analysis of the street network and the intervention areas.

Finally, we present findings from our investigation into the spatial variation of intersection density across the municipality. To achieve this, we employed two visualization methods that allow for a comprehensive assessment of intersection distribution, namely the hexbin plot and the kernel density estimation (KDE) plot. The first one, by using color intensity, represents the number of intersections within each bin (areas shaped like hexagons) (see Fig. 4.16). Regions with a higher density of intersections are clearly demarcated, revealing patterns and clusters of connectivity within the urban grid. The second plot, provides an estimation of the intersection density as a smooth, continuous surface. Through this method, we are able to discern a more nuanced view of how intersections are distributed in relation to the street layout, which is particularly useful in understanding the gradient of connectivity throughout the area (see Fig. 4.17). Regarding the hexbin plot, the color scale ranges from purple (representing lower intersection density) to yellow (indicating higher intersection density), with a threshold set to a maximum value (yellow indicates hexagons with 20 or more intersections). High-density areas, likely the urban center or other areas of intense activity, are marked by warmer colors, showing a higher number of intersections. Conversely, cooler colors signify regions with fewer intersections, which could be residential zones or peripheral areas of the city. Regarding the KDE Plot, the intensity of the color signifies also here the estimated density of intersections, with warmer colors representing higher densities. Not surprisingly, the highest density areas align with the centre of the city where also the major transit routes occur. Together, these two plots complement each other by providing both a granular view (hexbin) and a smoothed overview (KDE) of intersection density. These findings are particularly relevant when considering the bikeability and walkability of different regions within the municipality, as areas with a greater density of intersections typically correlate with enhanced accessibility and navigability for pedestrians and cyclists [177]. Consequently, the analysis not only maps out the current state of intersection density but also provides crucial data that can inform urban planning and the development of sustainable transportation infrastructure.

Regarding Bolzano, the result, which can be seen here (Fig. 4.18), highlight that:

- Degree Centrality: on the top left is visible the graph representing the nodes of the city colored by their degree centrality. The central area of the graph is also here densely populated with nodes, and the colors here are predominantly between 4 and 7, indicating a moderate degree of centrality. There are a just a few clusters of nodes with lower centrality values, indicated by their cooler colors. Except for the south east part, which has very low values, the overall structure of the graph seems to be also here organic.
- Closeness Centrality: on the top right is visible the graph representing the nodes of Bolzano network colored by their closeness centrality. The distribution of nodes shows a clear gradient from the center outwards, with nodes in the center of the network in yellow-green. The central and the surrounding areas exhibit high closeness centrality. Intermediate

values, represented by shades of green and light blue, radiate outward from the central cluster, indicating a gradient of decreasing centrality. This pattern suggests that there is a moderate level of accessibility extending from the center towards the outer areas. The nodes on the periphery of the network are colored in dark purple, marking them as the least central. These areas exhibit the highest average path lengths within the network, signifying that they are the least accessible in terms of direct paths to other nodes. These areas may represent the outskirts or less developed regions of Bolzano. In conclusion, the distribution of centrality values presents a clear radial decay pattern, where the highest values are concentrated in the core of the network and gradually diminish towards the edges. This is typical of cities where the central areas are well-connected hubs, and the connectivity decreases as one moves towards the outskirts.

- **Betweenness Centrality:** On the bottom left is visible the graph that illustrates the betweenness centrality of the nodes within the city of Bolzano. Here, we observe that nodes with higher betweenness centrality are not as clustered as in the closeness centrality visualization. Instead, high-centrality nodes (in lighter colors) are scattered across the network, indicating that these nodes are critical in connecting various parts of the city. Nodes with the highest betweenness centrality, shown in yellow, seem to be found on major streets and intersections that serve as key connectors within the network. The overall distribution of betweenness centrality appears to be less centralized than closeness centrality. The high-betweenness nodes form a kind of spine or arterial paths through the city, highlighting the main routes that channel movement. Meanwhile, the majority of nodes, colored in darker shades, suggest streets and intersections that are less critical to overall connectivity but may serve local neighborhoods or less trafficked areas. The majority of nodes are purple, suggesting that many nodes have a lower betweenness centrality, playing a less critical role as intermediaries in the network's connectivity.
- **Edge Betweenness Centrality:** displayed in the lower right corner is a graphical representation of edge betweenness centrality, with a spectrum spanning from a nadir of 0.000 to an apex just shy of 0.030. This visual complements the node betweenness centrality graph, corroborating the observed distribution patterns. Predominantly, the edges exhibit minimal betweenness centrality, indicating a lower likelihood of their traversing the shortest paths within the network. However, a concentration of streets in the central region registers heightened betweenness centrality. This phenomenon suggests that these particular streets may bear a greater volume of traffic, serving as pivotal conduits within Bolzano's urban grid, thus enhancing overall network connectivity.

Similar to the approach taken with Trento, centrality metrics for Bolzano were recalculated incorporating level of traffic stress (LTS) values as weights. When examining the resulting graph (see Fig. 4.19), the visual outcome is largely consistent with that of the initial calculations. However, a notable divergence is observed within the Closeness Centrality metric; it displays a shift towards greener hues in the central regions of the city, indicating a modification in centrality due to the inclusion of traffic stress considerations.

The following outlines the descriptive statistical analysis of the centrality metrics distribution:

- **Degree Centrality:** on the top of Fig. 4.20, is depicted a histogram overlaid with a kernel density estimation (KDE) curve that represents the distribution of degree centrality values in the network. The tallest bar is near in this case to 12. On average, a node in this network has a degree centrality value of approximately 10.45. The distribution appears to be less right-skewed than Trento. The majority of bars are more present on the right side,

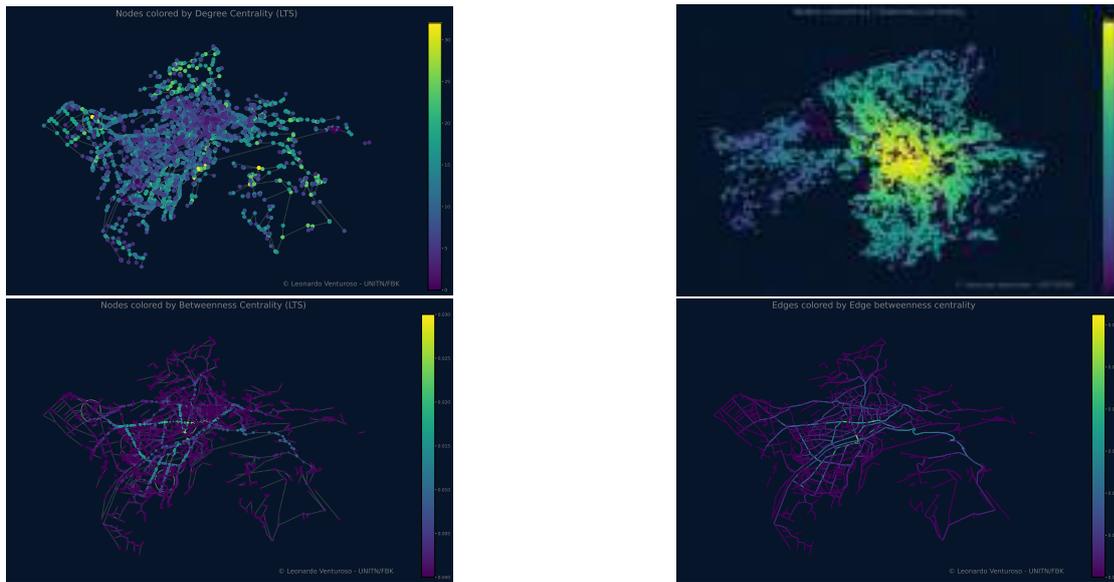


Figure 4.18: Top Left: Node Degree Centrality; Top Right: Node Closeness Centrality; Bottom Left: Node Betweenness Centrality; Bottom Right: Edge Betweenness Centrality — all within the Bolzano Network.

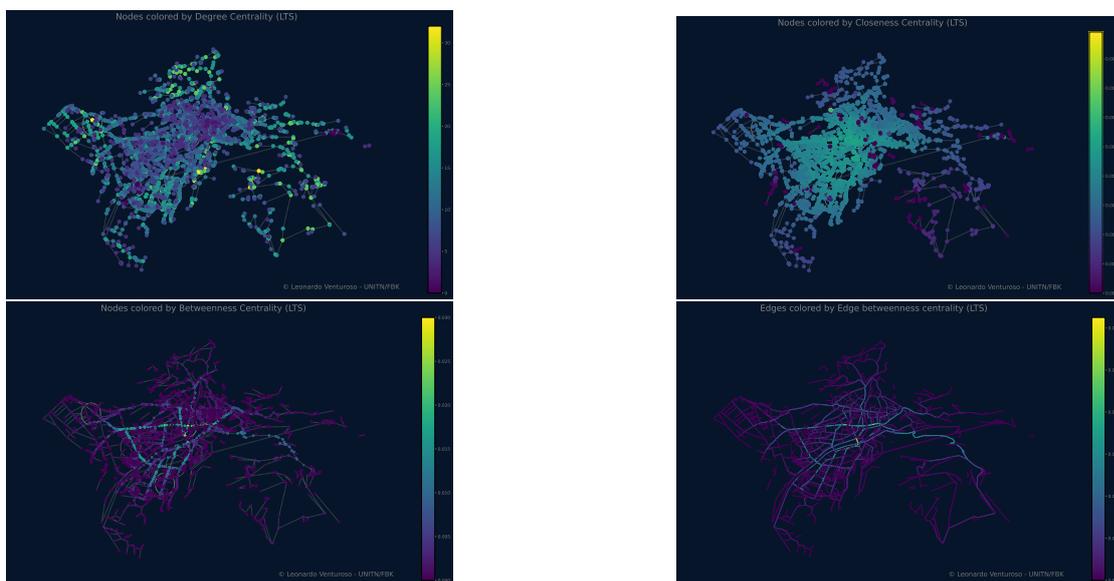


Figure 4.19: Top Left: LTS - Node Degree Centrality; Top Right: LTS - Node Closeness Centrality; Bottom Left: LTS - Node Betweenness Centrality; Bottom Right: LTS - Edge Betweenness Centrality — all within the Bolzano Network.

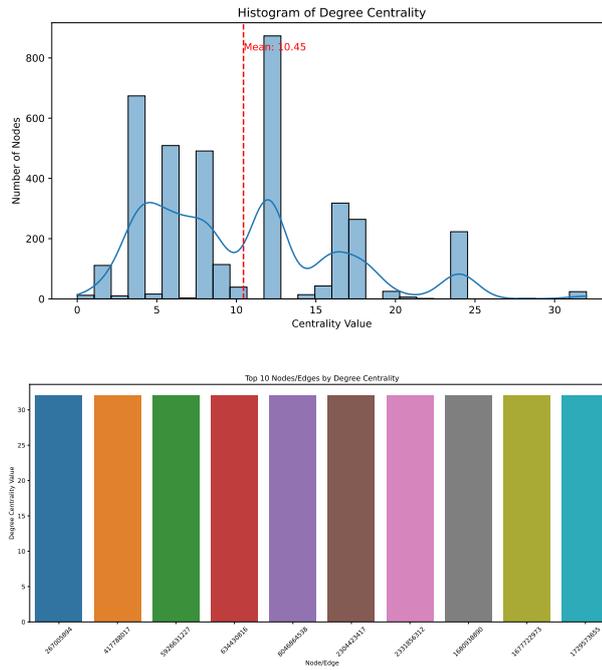


Figure 4.20: Top: Distribution of nodes for Degree Centrality values for Bolzano. Bottom: Top 10 nodes for Degree Centrality values in Bolzano.

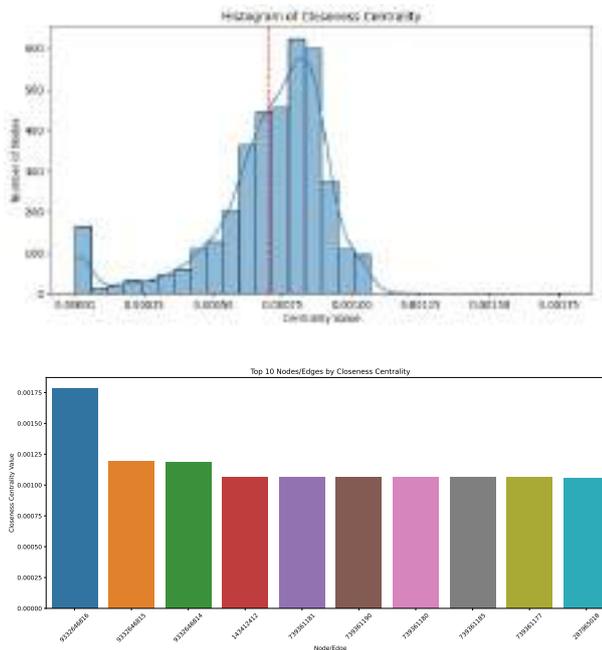


Figure 4.21: Top: Distribution of nodes for Closeness Centrality values for Bolzano. Bottom: Top 10 nodes for Closeness Centrality values in Bolzano.

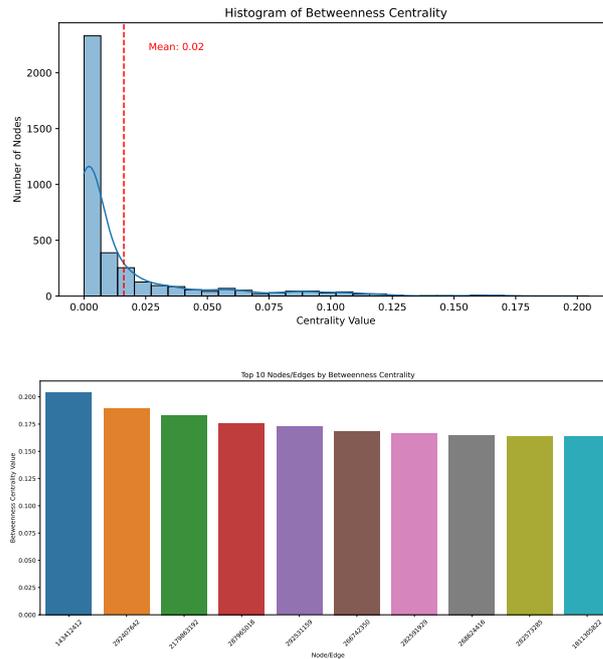


Figure 4.22: Top: Distribution of nodes for Betweenness Centrality values for Bolzano. Bottom: Top 10 nodes for Betweenness Centrality values in Bolzano.

between 0 and 12. On the bottom of the figure we can see the top 10 nodes for degree centrality values. The highest values is around 32 nodes.

- **Closeness Centrality:** on the top of Fig. 4.21, is depicted a similar histogram overlaid with the kernel density estimation curve. Looking at the shape of the distribution, it appears to be approximately normal (Gaussian), with most of the nodes exhibiting centrality values around a central peak, which suggests that most nodes have a moderate level of centrality within the network. The highest frequency of nodes falls within the bin around the red dashed line. There is a large spike indicates that a significant number of nodes have a closeness centrality close to 0.0008. The mean of the distribution is close to the mode of the histogram, indicating that the distribution is a bit symmetric. There are fewer nodes with very low or very high closeness centrality values, as indicated by the lower bars at the extremes of the x-axis. This is typical in a well-connected urban network, where most intersections are reasonably well integrated into the street layout, with only a few being very central or very peripheral.
- **Betweenness Centrality:** at the upper section of Fig. 4.22, a histogram presents the betweenness centrality values across the Bolzano network. This histogram demonstrates a rightward skew, characterized by a precipitous decline in frequency as centrality values ascend. With the mean centrality marked at 0.02, the distribution exhibits a pronounced skew towards the lower end. The count of nodes rapidly diminishes in conjunction with rising betweenness centrality, suggesting that only a select few nodes hold substantial sway over the network's throughput. This distribution pattern signifies that while most nodes do not play a pivotal role in the network's connectivity, a handful of nodes are integral to

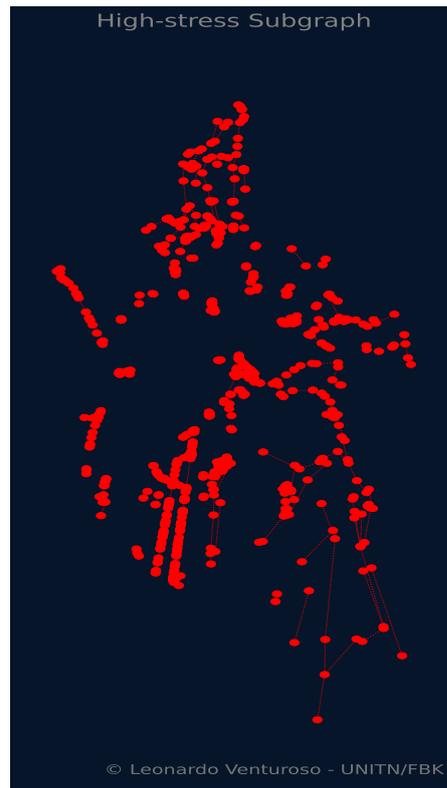


Figure 4.23: High Stress Subgraph in Bolzano

the majority of the shortest paths.

The areas subjected to high stress are illustrated in the Fig. 4.23). The nodes experiencing high levels of stress predominantly cluster in the northern and southern areas of the city, in contrast to the western side and the central region, where they appear more dispersed. This pattern may be linked to the denser mountainous regions to the north and south, which likely contribute to a lower degree of node connectivity. The top five components characterized by high stress have been identified and are presented herein. (Fig. 4.24).

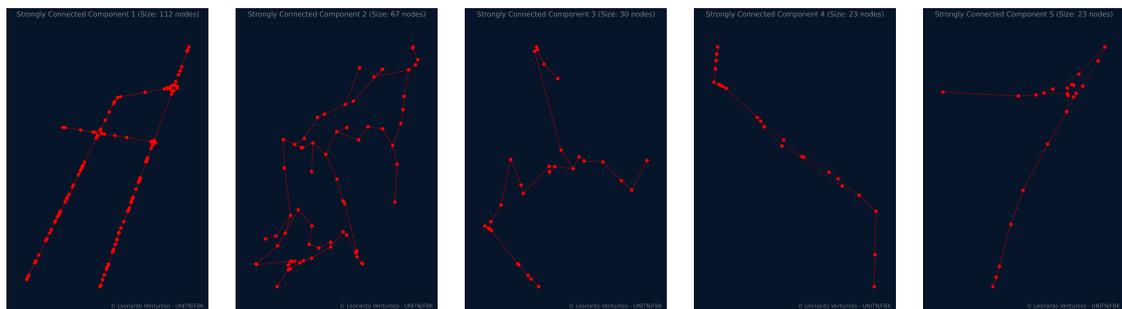


Figure 4.24: Top 5 High-Stress Connected Components in Bolzano

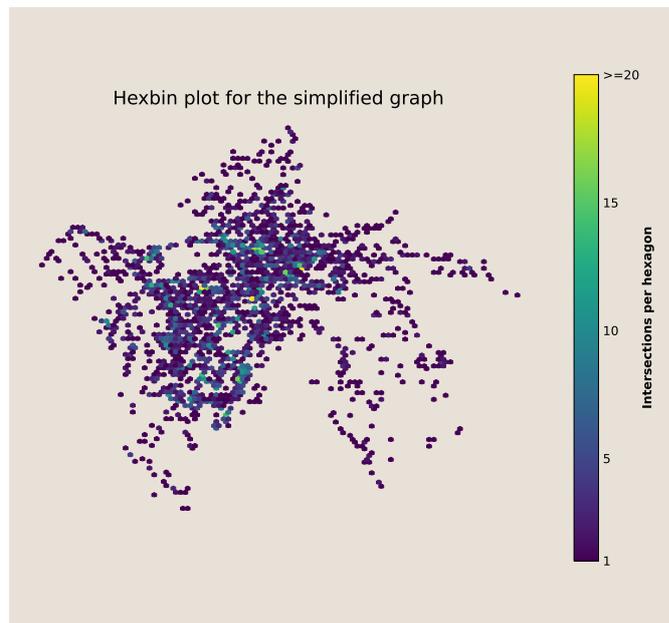


Figure 4.25: Hexbin map of Bolzano - Intersection Density

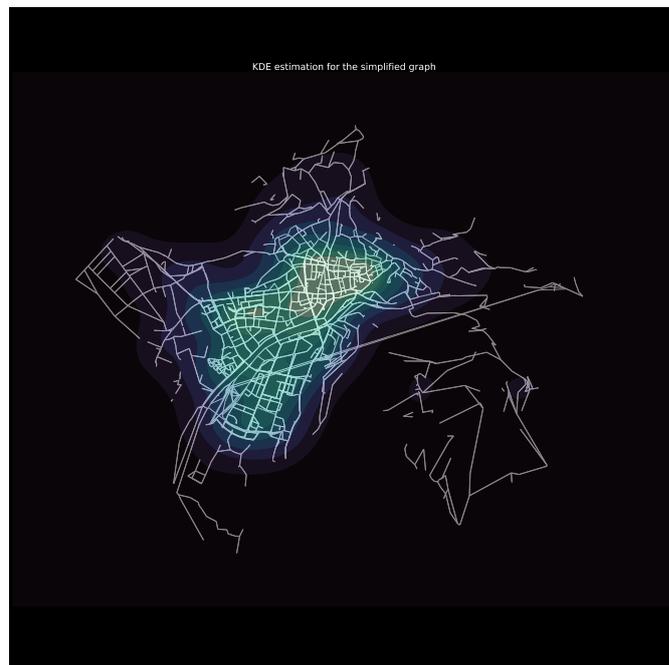


Figure 4.26: Kernel Density Estimation Map of Bolzano - Intersection Density

Regarding the intersection density across Bolzano these are the two plots. The hexbin plot represents the number of intersections within each bin of the city (see Fig. 4.25). The graph does not display distinct clusters of intersections, indicating that overall intersection density is not so high. The accompanying Kernel Density Estimation Map offers a more discernible representation of intersection density. However, without the context provided by the first graph, it would be challenging to ascertain the generally low intersection density and the fact that the clusters do not correspond to areas with a general high degree of intersections.(see Fig. 4.26).

Cluster Analysis

As mentioned in the Methodology section, a comprehensive cluster analysis is implemented to identify and visualize high-stress zones within urban road networks.

The cluster analysis conducted on Trento's street network using the DBSCAN algorithm reveals distinct groupings of high traffic stress areas, as visualized in the provided graphs (see Fig.4.27). The plots showcase the spatial concentrations of traffic stress at various intersections throughout the city. The first graph focuses solely on the geographic coordinates. The clusters here are more pronounced against the plain background, allowing for an unobstructed analysis of the spatial relationships. The arrangement of clusters across the latitudinal and longitudinal planes highlights concentrated areas of traffic stress within Trento, particularly noticeable in the central right-hand section of the mapped area. The second graph presents a layered view over the city's map, with clusters of intersections depicted in a spectrum of colors corresponding to different clusters. These clusters are scattered across the city's layout, showing that the densely concentrated area identified earlier is the one close to the center of the city. This suggests a region where traffic stress is particularly high, potentially indicating bottlenecks or areas that lack sufficient infrastructure to accommodate safe and accessible bicycle traffic. Quantitative validation of the clusters' significance comes from the evaluation metrics. The Silhouette Score of 0.703 suggests that the clusters are appropriately compact and well-defined, meaning that intersections within a cluster are close to each other while being distant from other clusters. The Davies-Bouldin Score, at a low of 0.318, supports this by indicating that the clusters are well-separated and distinct. Finally, the high Calinski-Harabasz Score of 15375.52 confirms that the clusters are indeed meaningful, showing that they are tight-knit groups well-separated from each other.

The cluster analysis conducted with the HDBSCAN algorithm on the street network of Trento, specifically focusing on intersections characterized by high levels of traffic stress. This is an advanced method particularly effective for identifying clusters of varying densities.

According to the results (see Fig.4.28), we see in the first graph a geographic scatter plot of the intersections, with each point color-coded according to the cluster it belongs to. The clusters are spread across the map, but also here there's a notable concentration of clustered points towards the center. This dense aggregation, particularly of warm-colored points, indicates regions within Trento that experience higher traffic stress.

The second graph, as with DBSCAN, overlays the clusters on an actual map of Trento, providing context to the geographic data. This visualization allows us to see the exact urban layout and how the high-stress intersections relate to the city's streets and neighborhoods.

In this case, the Silhouette Score of 0.555 indicates a moderate level of definition for the clusters, although is lower compared to the DBSCAN. This means that while there is some overlap, the clusters generally represent distinct groupings of high traffic stress intersections. The Davies-Bouldin Score of 0.544 suggests that the clusters are reasonably compact and well-separated, which is desirable for clear cluster delineation. Finally, the Calinski-Harabasz Score of 5826.095 reflects that the clusters are quite distinct and well-defined relative to each other,

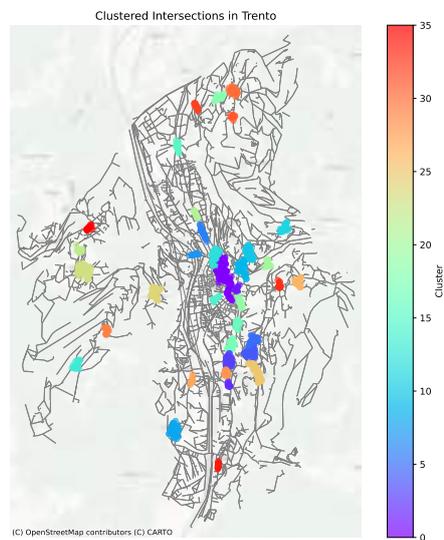
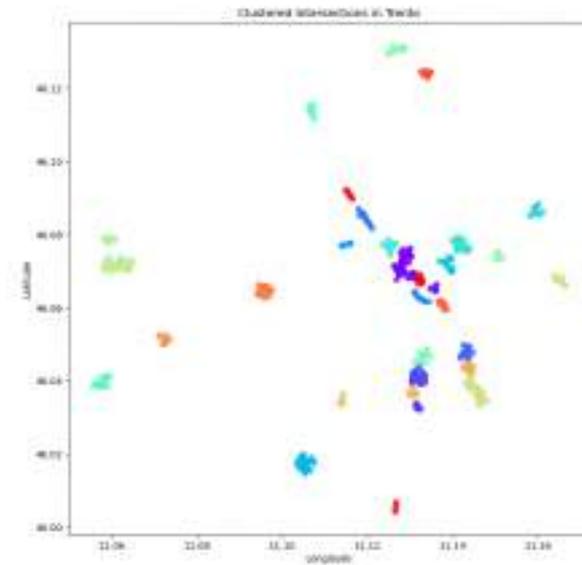


Figure 4.27: Top: DBSCAN Clustering Method applied for Trento. Bottom: DBSCAN Clustering Method applied for Trento over city network layer.

although this score is lower compared to the DBSCAN method previously mentioned, indicating differences in cluster cohesion and separation.

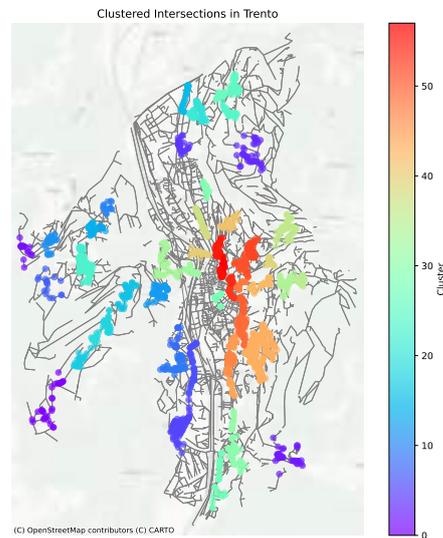
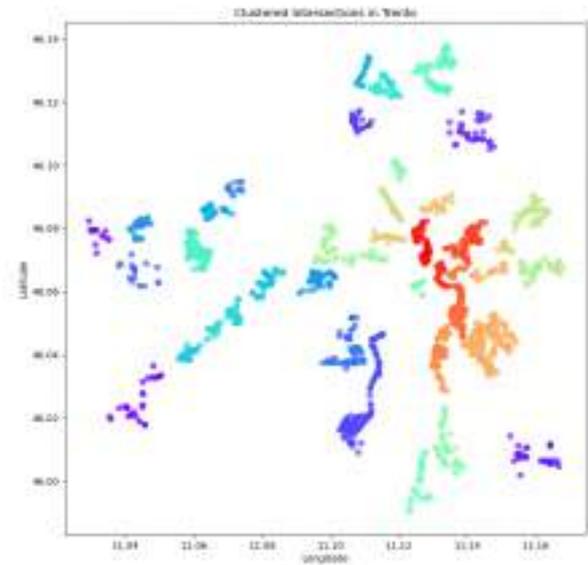


Figure 4.28: Top: HDBSCAN Clustering Method applied for Trento. Bottom: HDBSCAN Clustering Method applied for Trento over city network layer.

Regarding the last clustering method, which is the OPTICS, this aims to identify high-stress traffic areas in Trento, based on the intersections' high traffic stress levels. This method is adept at detecting clusters of varying density, which is particularly useful for spatial data that do not have a uniform distribution.

In the first graph, we can observe that some clusters appear to be closely grouped, such as the red and blue clusters towards the lower left of the plot, which indicates a localized area of high traffic stress. Other clusters are more dispersed, suggesting areas where high traffic stress is present but not as densely packed.

The second graph reveals that the clusters are not uniformly distributed throughout the city; instead, they form notable groupings. For instance, certain clusters appear to follow major thoroughfares.

Both graphs (see Fig. 4.29) highlight specific regions within Trento that may require urban planning interventions. These could include areas in the central part of the city where clusters are denser, potentially correlating with higher volumes of traffic or less safe and accessible bike facilities. Conversely, the presence of more isolated clusters on the outskirts might point to individual intersections that are problematic for other reasons that must be investigated (the areas on the lower right regards a mountain areas).

The cluster analysis is supported by the metrics. Specifically, the Silhouette Score of 0.733 indicates strong cluster definition, the Davies-Bouldin Score of 0.339, which is relatively low, suggests that the clusters are well-separated and compact and finally, the Calinski-Harabasz Score of 9319.336 signifies that the clusters are distinct and well-defined. Although different from the previous scores obtained with DBSCAN and HDBSCAN, this score indicates clear delineation and separation of clusters.

Overall, each algorithm has its strengths as shown by these scores: OPTICS has the highest Silhouette Score, indicating the best cluster cohesion and separation. DBSCAN has the lowest Davies-Bouldin Score and the highest Calinski-Harabasz Score, suggesting the best overall cluster separation and density. HDBSCAN scores are lower than the other two algorithms, indicating that while its clusters are reasonable, they are not as compact or separated as those identified by OPTICS or DBSCAN. In conclusion, the ultimate choice leans toward the DBSCAN algorithm, as it demonstrates superior performance with respect to both the Davies-Bouldin score and the Calinski index and also from a visual point of view.

Regarding Bolzano, the use of DBSCAN algorithm reveals distinct groupings of high traffic stress areas, as visualized in the provided graphs (see Fig.4.30). As it can be seen, traffic stress concentrations are markedly evident in the city's southern region. A Silhouette Score of 0.652 indicates that the clusters are comparatively well-defined and distinguishable from one another. This is corroborated by the Davies-Bouldin Score, which stands at a low 0.390, suggesting clear separation and distinctiveness of the clusters. Furthermore, a substantial Calinski-Harabasz Score of 1680.83 reinforces the meaningfulness of these clusters.

The HDBSCAN algorithm is also performed on the street network of Bolzano. According to the results (see Fig.4.31), we see points of the clusters spread across the map highlighting different areas of concentration. The only area without any high stress nodes seems to be the one regarding the center of the city. The Silhouette Score of 0.442 indicate a moderate level of definition, lower compared to the previous method. The Davies-Bouldin of 0.580 suggest that the clusters are moderately compact. A Calinski-Harabasz Score of 638.45 is moderate, indicating some level of cluster density and separation but not particularly high.

Regarding the OPTICS clustering method, we can observe that some clusters appear to be closely grouped in the southern and in the eastern side. Other clusters are more dispersed (example: the one in the north). The map highlights that the clusters are not uniformly distributed throughout the city (see Fig. 4.32). Differently from the other two methods a cluster is found

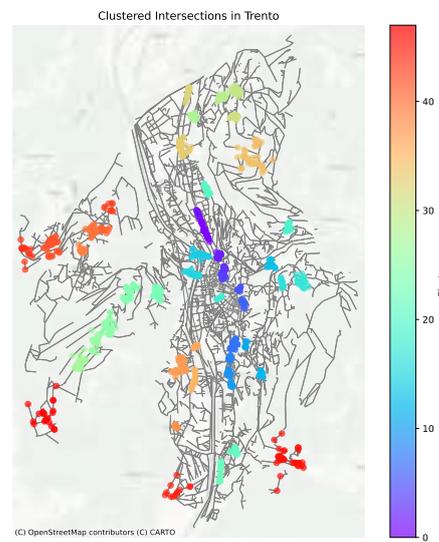
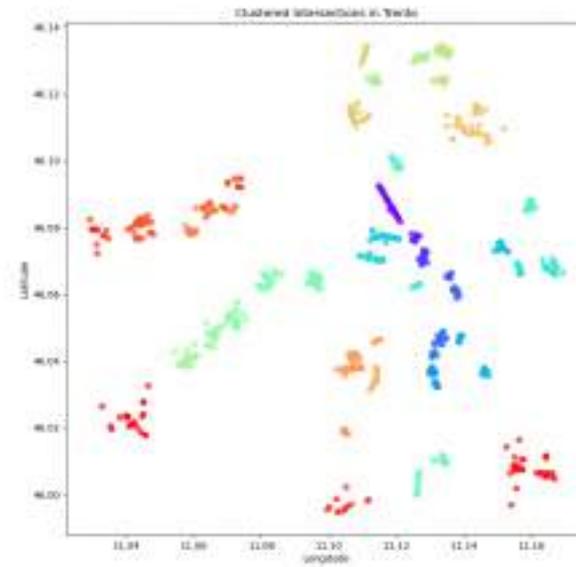


Figure 4.29: Top: OPTICS Clustering Method applied for Trento. Bottom: OPTICS Clustering Method applied for Trento over city network layer.

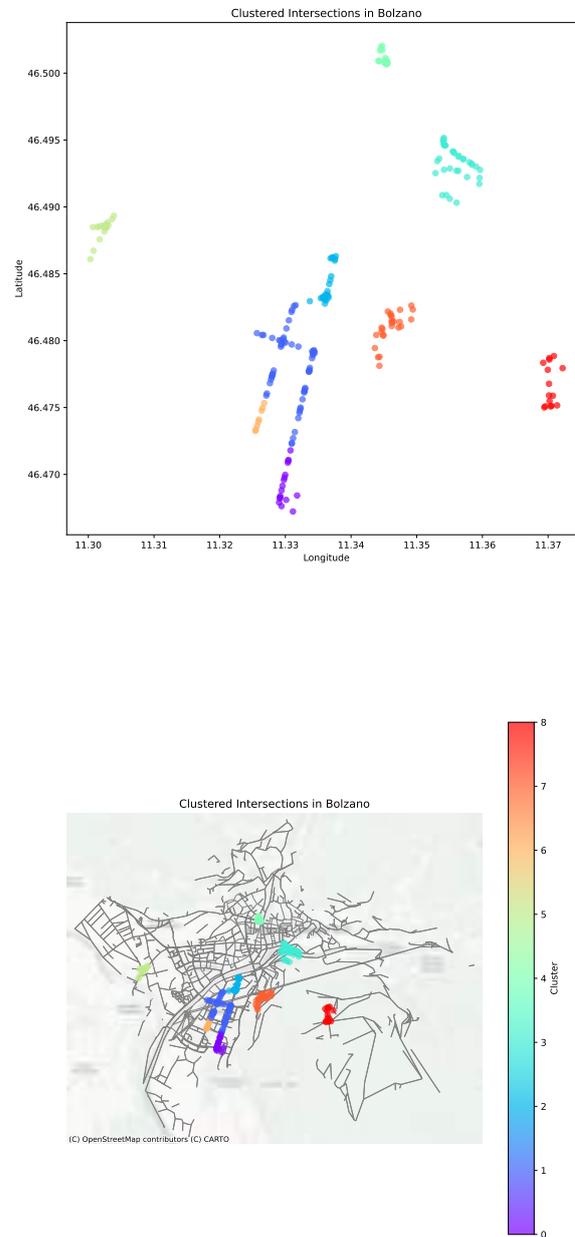


Figure 4.30: Top: DBSCAN Clustering Method applied for Bolzano. Bottom: DBSCAN Clustering Method applied for Bolzano over city network layer.

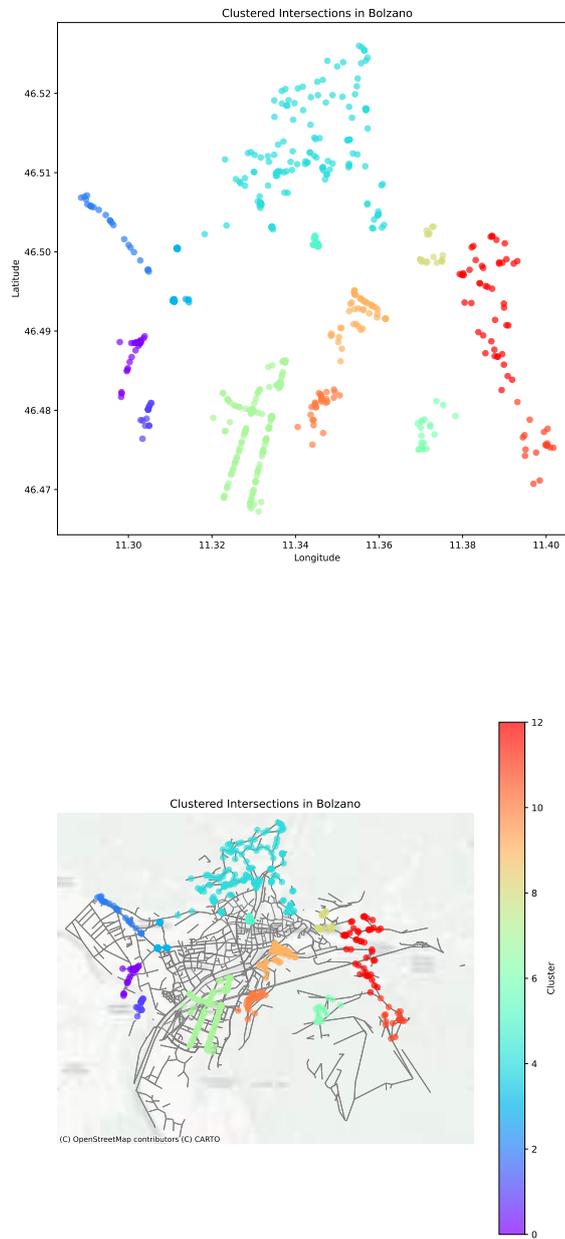


Figure 4.31: Top: HDBSCAN Clustering Method applied for Bolzano. Bottom: HDBSCAN Clustering Method applied for Bolzano over city network layer.

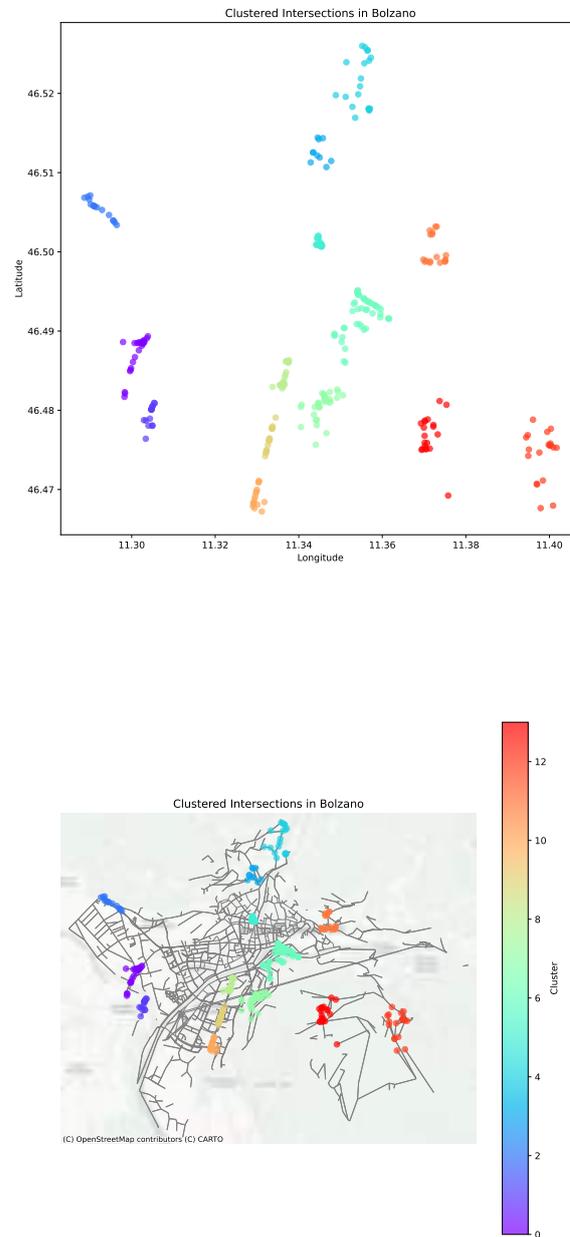


Figure 4.32: Top: OPTICS Clustering Method applied for Bolzano. Bottom: OPTICS Clustering Method applied for Bolzano over city network layer.

in the center of the city. A Silhouette Score of 0.723 reflects robust cluster definition, while the relatively low Davies-Bouldin Score of 0.348 implies that the clusters are both well-separated and compact. Furthermore, the Calinski-Harabasz Score of 2371.376 denotes that the clusters are distinct and well-defined. These scores suggest a pronounced delineation and separation among the clusters. Overall, all three methods demonstrate effective clustering capabilities, with OPTICS showing the strongest performance across all metrics. HDBSCAN and DBSCAN also perform well, with HDBSCAN being more moderate in its clustering effectiveness compared to the other two methods.

High Stress Gap Analysis

Here we present the outcomes of applying the adapted LTS - IPDC (Identify, Prioritize, Decluster, Classify) procedure to the urban networks of Trento and Bolzano, with a specific focus on existing infrastructure where bicycle can move and its interplay with levels of traffic stress. This approach scrutinizes the network to identify high-stress traffic areas that are barriers to low-stress subgraphs, effectively segmenting the network into more and less bike-friendly zones. We will begin by presenting the results for each individual step for Trento. Ultimately, we will also report and discuss the outcomes for Bolzano.

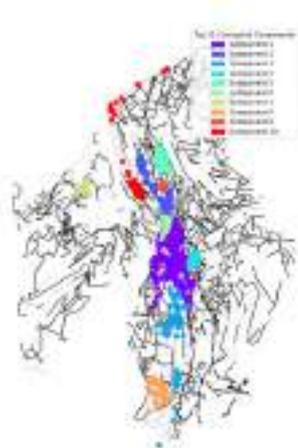


Figure 4.33: Top 10 Low Stress Connected Components in Trento

The first step, consist in identifying the connected components of our network, which are clusters of interconnected low-stress paths. The top 10 largest connected components are then visualized (see Fig 4.33). This step is crucial for a first visual analysis and identification gaps in the network. As it can be seen, the first two (the ones in purple/blue) are the ones in the center of the city and are also the biggest ones. They are followed by the the light blue and the teal components which together occupy the all center of Trento. We infer that not surprisingly, the areas where traffic is often restricted are those where we can witness low-stress networks. Then, we proceed by identifying segments of the road network that exhibit high levels of traffic stress (namely, LTS 3 or 4) and the ones with low levels of traffic stress (namely, LTS 1 or 2) (see Fig. 4.34). Then we identify the '*gaps*' which are segments of high traffic stress that separate low



Figure 4.34: High-Low Stress components in Trento



Figure 4.35: Gaps plot in Trento

stress subgraphs (see Fig. 4.35) and in particular can be defined as high stress links between low stress contact nodes (see Fig. 4.36). Finally, the list of gaps is filtered by a detour factor of 2 meters. These segments were marked as potential barriers to cycling, deterring cyclists who prefer low-stress routes. The procedure then prioritized these high-stress segments based on factors such as the link betweenness connectivity that we mentioned in the Methodology section. Then, based on this measure, we ranked the gaps in terms of benefit metric and we obtain a list of paths and its corresponding benefit metric. The first 10 gaps can be seen in the Table 4.3.

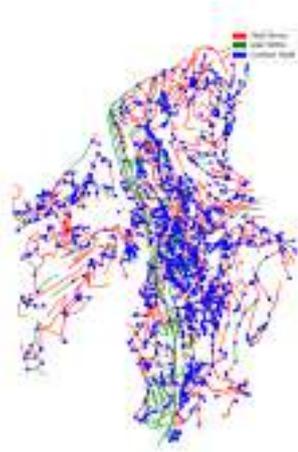


Figure 4.36: Low Stress Contact Nodes in Trento

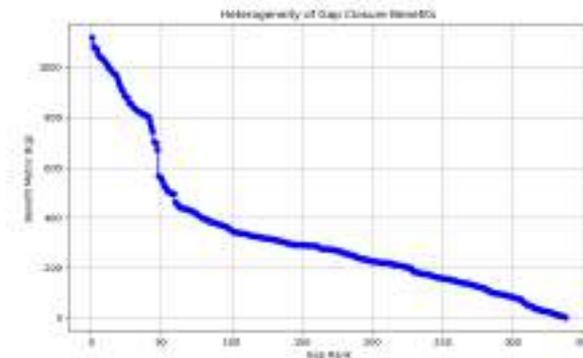


Figure 4.37: Benefit Metric and Gap Rank graph in Trento

We plot the result as it can be seen here (Fig. 4.37). Then, in the declustering step, we first filter out gaps with a benefit metric below the threshold using the graph 4.37 and analyzed the spatial distribution of high-stress gaps to separate out clusters of contiguous high-stress roads starting from an initial list of 327 gaps. The final result is a list of 8 declustered gaps that can be seen here in the top image (Fig. 4.38). Finally, the road segments were classified according to their type (e.g street or bridge). The result can be seen in the image at the bottom of 4.38. According to the final list there are 8 most important high stress gaps for Trento. The first is the one of the section that includes the Ravina Bridge (Ponte's Street), the traffic circle and Al Desert Street, and the traffic circle between Hubert Jedin Street and Alcide De Gasperi Street, an area that is notoriously busy at certain times (early morning and late afternoon). The second concerns another rather busy area of South Trento, namely between the traffic circle at Verona Avenue/Madonna Bianca Street and the traffic circle between Enrico Fermi Street and Verona

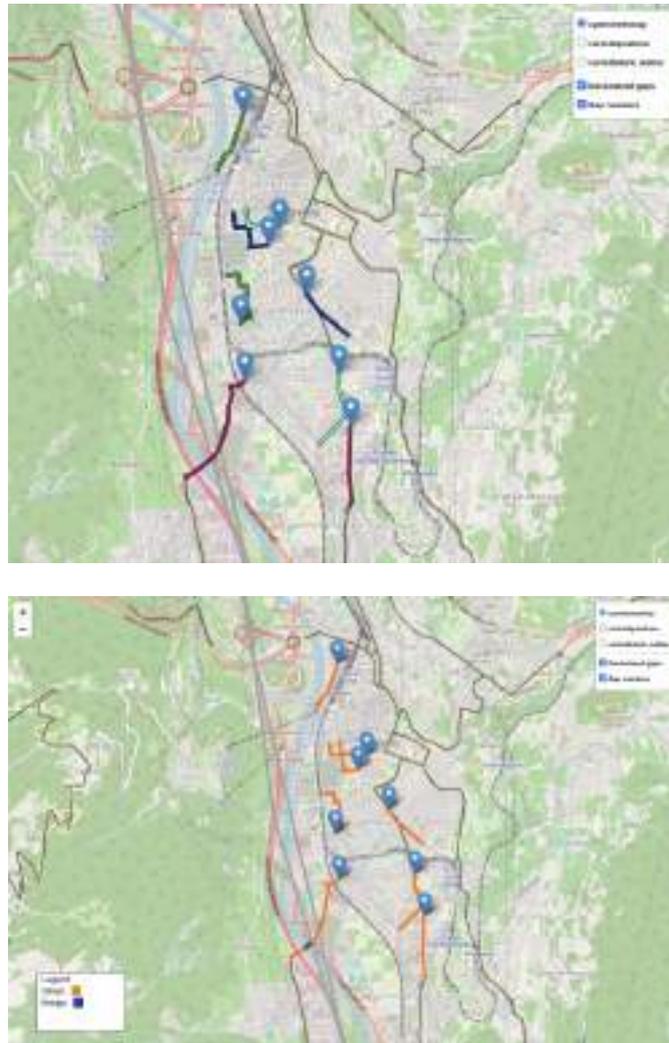


Figure 4.38: Top: Declustered Gaps in Trento. Bottom: Classified Gaps in Trento

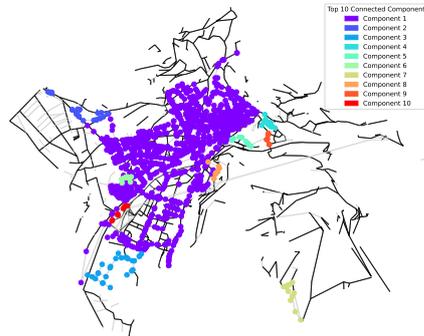


Figure 4.39: Top 10 Low Stress Connected Components in Bolzano

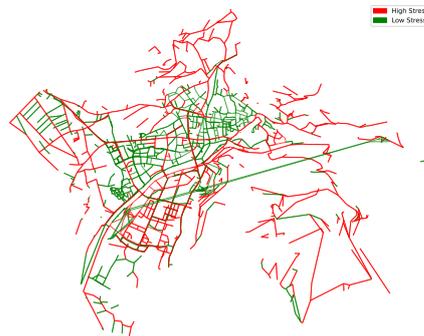


Figure 4.40: High-Low Stress components in Bolzano

Avenue. The third one is very close to the second and is the one between Padre Eusebio Chini Street, the traffic circle mentioned above and another traffic circle at Enrico Fermi Street, Alcide De Gasperi Street. Also in this case we know that this is a very busy area at certain times of the day. The fourth gap is always located between a traffic circle, namely the one between Via Vittorio Veneto, Via Dei Muredei and Via Giuseppe Giusti in the San Pio X area, and the section that goes back over Via Giuseppe Giusti to a side road near the Celestino Endrici Archbishop's College. The fifth gap is the one that starts from the intersection of Milano Street

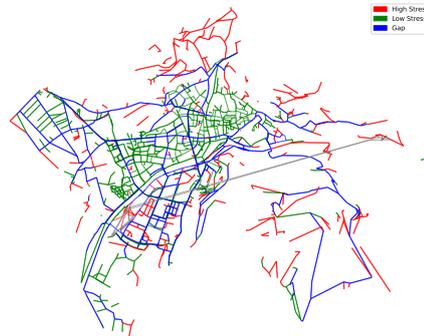


Figure 4.41: Gaps plot in Bolzano

and 3 Novembre Street, passes through the Cavalleggeri Bridge, Verona Avenue, Gocciadoro Street and the traffic circle with Paolo Orsi Street in the Santa Chiara Hospital area. The sixth and seventh gaps pertain to a highly central zone, which corresponds to the area spanning Giuseppe Verdi Street, Antonio Rosmini Street, and Dei Travai Street, extending to Fiera Square. Finally the last one starts in the Trento/Sardagna cable car area near the traffic circle, passing along Lungadige Giacomo Leopardi and Michelangelo Buonarroti Street to the intersection with Dosso Dossi Street, another busy area at certain times of the day.

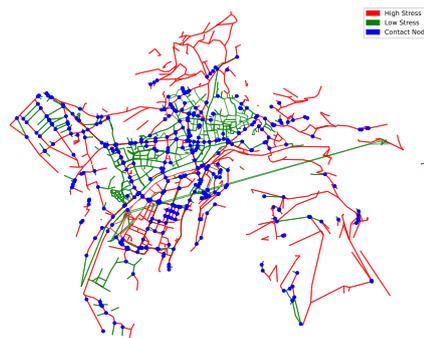


Figure 4.42: Low Stress Contact Nodes in Bolzano

Gap From	Gap To	Benefit Metric
4085216047	248815336	1118.959
3098727651	248815336	1084.367
4085216047	258070520	1076.685
9205502	248815336	1073.241
4085216047	941909659	1051.848
3098727651	258070520	1042.093
3098187788	248815336	1039.848
1635214896	248815336	1036.193
9205502	258070520	1030.967
673248086	248815336	1022.288

Table 4.3: Benefit Metric for Identified Gaps

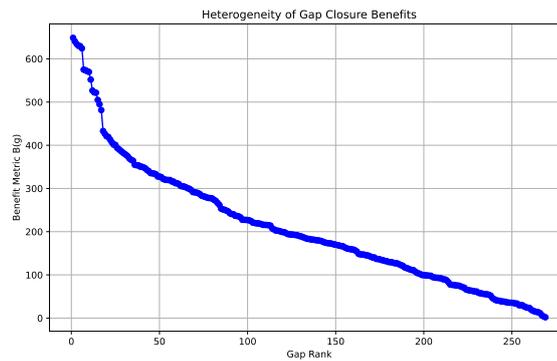


Figure 4.43: Benefit Metric and Gap Rank graph in Bolzano

Regarding Bolzano, the top 10 largest connected components are then visualized (see Fig 4.39). In the center of the city, there's a prominent purple area, significantly larger than its surroundings. Adjacent to it are smaller blue and light blue areas. Consistent with previous case, regions near the city's center exhibit the lowest levels of traffic stress. In Fig.4.40, the road network segments with high traffic stress (specifically, LTS 3 or 4) are depicted alongside those with low traffic stress (specifically, LTS 1 or 2). Following this, *gaps* in the network are presented in Fig. 4.41, and subsequently, the visualization of relatively low-stress contact nodes is illustrated in Fig. 4.42. Based on link betweenness connectivity gaps are then ordered in terms of benefit metric. The first 10 can be seen in the Table 4.4.

We plot the result as it can be see here (Fig. 4.43). Then, gaps with a benefit metric below the threshold are filtered using the graph 4.43 and the spatial distribution of high-stress gaps is analyzed to separate out clusters of contiguous high-stress roads starting from an initial list of 269 gaps. The final result is a list of 3 declustered and classified gaps that can be seen here

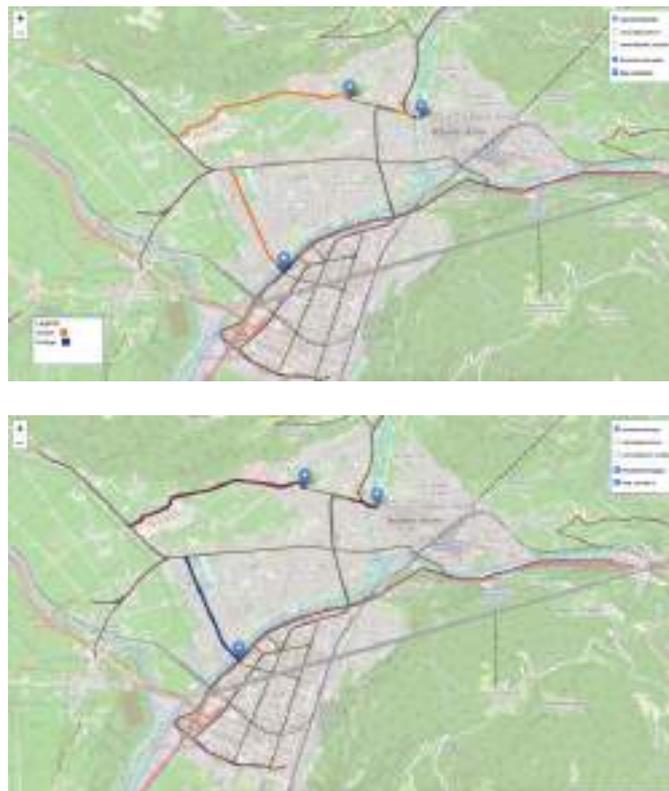


Figure 4.44: Top: Declassified Gaps in Bolzano. Bottom: Classified Gaps in Bolzano

Gap From	Gap To	Benefit Metric
282800486	10055891648	648.604
10055891657	282800486	640.538
10055296928	282800486	634.657
266484972	282800486	630.219
282800486	10057800586	629.535
2430291015	771200575	624.317
282800486	646601506	575.062
282800486	10057733973	573.687
2282909209	282800486	571.443
10057733983	282800486	569.787

Table 4.4: Benefit Metric for Identified Gaps

in the image (Fig. 4.44). The analysis identifies three critical high-stress gaps within Bolzano’s traffic network. The first notable gap spans the area from Resia Bridge along Resia Street to Malles Street. This segment forms a vital connection between the Don Bosco and Oltrisarco regions, particularly the industrial zone. The second key gap involves the Talvera Bridge, a crucial link between the Centro Storico (historical center) and San Quirino areas. The affected stretch extends from Talvera Bridge through Vittoria Square and Libertà Street, reaching up to Corso Orazio. The final significant gap encompasses a broader area: starting from Libertà Street, it includes Gries Square, Vittorio Veneto Street, San Maurizio Street, and the vicinity of Lorenz Böhler Street near the hospital area, concluding at Merano Street.

4.1.3 Qualitative Analysis

Urban Space Quality Assessment

In the results section, we delve into the findings obtained from the Urban Space Quality Assessment procedure for Trento and Bolzano, which was tailored to enhance the cycling experience in urban areas. The assessment was informed by the methodology adapted from Novack, Wang, and Zipf [177], focusing on the accessibility of Points of Interest (POIs) and travel times from a cyclist’s perspective. The primary visual output was a scatter plot that color-coded nodes based on the biking time to the nearest POI in Trento and Bolzano, offering a preliminary glimpse at accessibility within the network (see Fig. 4.45 and Fig. 4.46).

As it can be seen, the nearest POI for a specific node (amenities like cafe or bar, shops like bakery or supermarket and leisures like fitness centre) is accessible in the center areas within 5 minutes in proximity of the centers. While the time increases in more peripheral areas. To deepen our insights, another visualization took into account travel times to not only the closest POI but also the 5rd and 10th nearest, thereby providing a layered perspective on amenity availability using hexbins. This provides a more comprehensive picture of urban accessibility. According with the result, (see Fig. 4.47 and see Fig 4.48), the top-right panel represents biking times to the nearest POI. The nodes are aggregated into hexagonal bins, and the color intensity indicates the average travel time from the nodes within each bin. This visualization shows

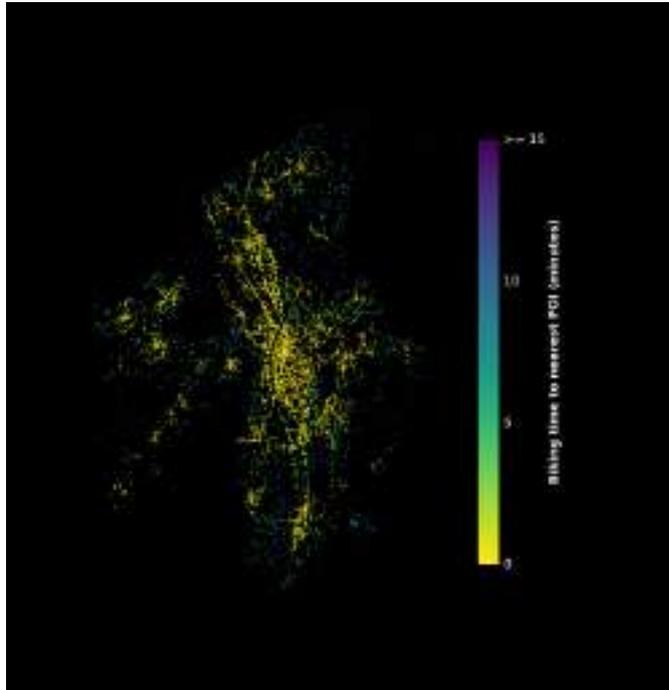


Figure 4.45: Biking time to the nearest POI within 15 minutes in Trento

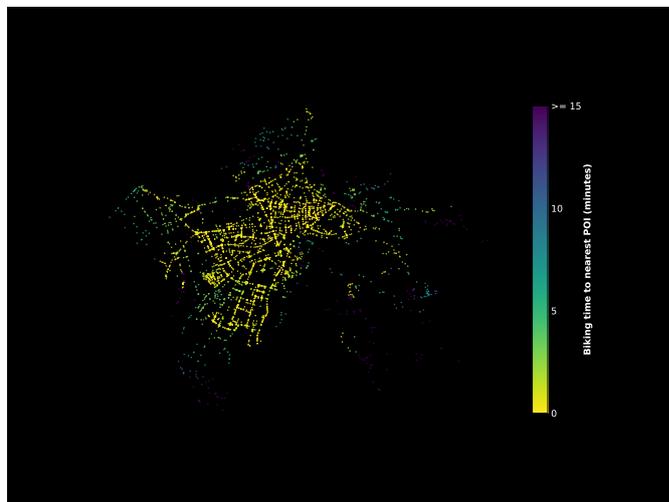


Figure 4.46: Biking time to the nearest POI within 15 minutes in Bolzano

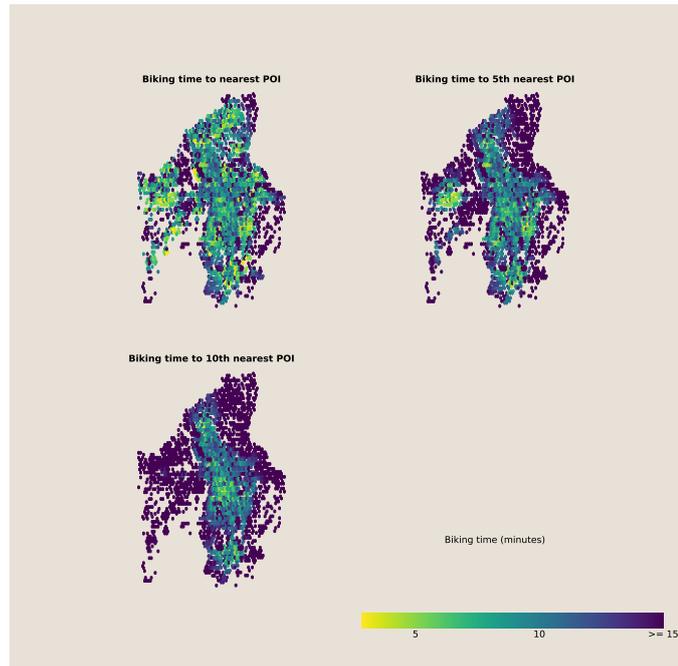


Figure 4.47: Biking time to the nearest POIs (5th and 10th) within 15 minutes in Trento

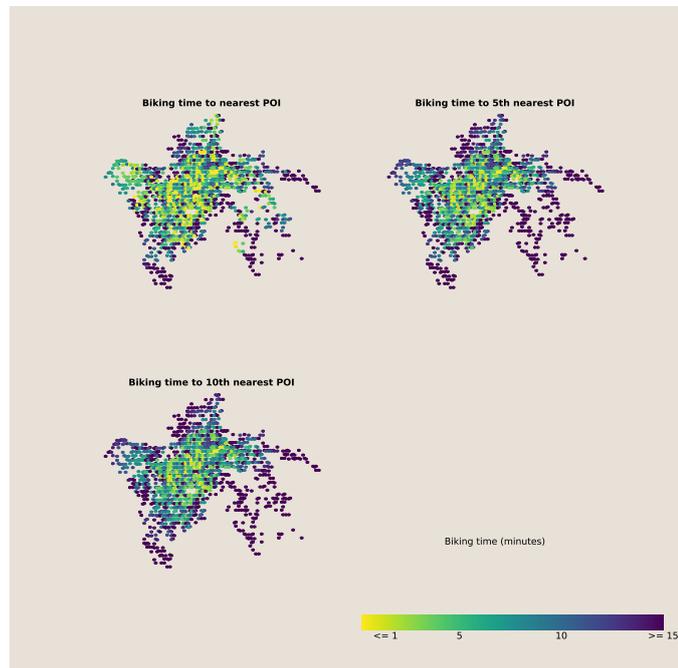


Figure 4.48: Biking time to the nearest POIs (5th and 10th) within 15 minutes in Bolzano

a concentration of shorter travel times in the urban centers where POIs are densely located, specifically for Bolzano. Areas with longer travel times might correspond to peripheral areas (industrial, mountains) where amenities are more sparse. The top-right panel extends the analysis to the 5th nearest POI. Compared to the nearest POI map, this panel may show a broader distribution of moderate travel times, reflecting areas where amenities are not as densely packed but still within a reasonable biking distance. Finally, the bottom panel focus on the biking time to the 10th nearest POI, indicating a significant increase in travel times. Times are shorter in urban center and longer travel distances are more common in peripheral areas (darker color ones). The hexbin step-by-step approach effectively declutters the visualization, allowing for the identification of patterns that are not immediately apparent in individual node-based analyses.

4.1.4 Bicycle Network Analysis (BNA)

In this section, we will present the results of the procedure used to calculate the Bicycle Network Analysis (BNA) score for the networks of Trento and Bolzano. These results will be showcased as maps that depict the grid of each municipality divided into H3 hexagons [104], each colored according to its BNA score, which reflect the level of bikeability and accessibility to common destination of that specific area. Alongside these maps, we will also display grids where each hexagon is colored based on the estimated number of inhabitants—a crucial piece of information needed for the computation. Regarding Trento, the city obtain an overall BNA score of 46.58. To calculate the score we combined the LTS analysis scores, the list of destinations 3.4 and the population data (see Fig. 4.49).

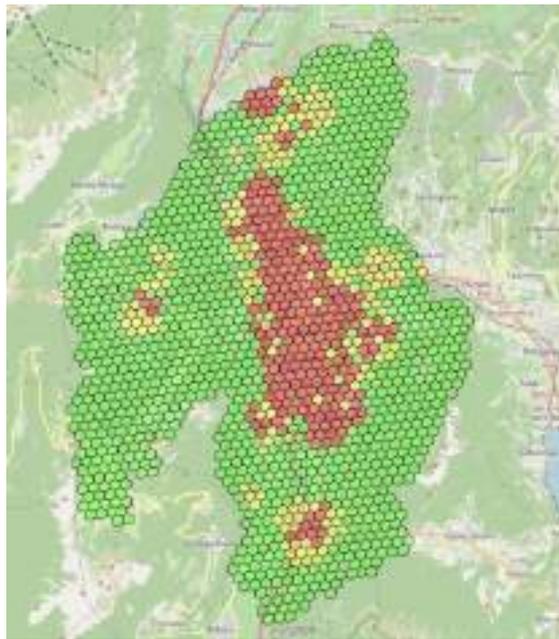


Figure 4.49: Population distribution in Trento H3 grid, resolution 200m. Red areas are the one more populated (more than 2% of population), followed by yellow one (more than 1% of population) and green one.

Each hexagon of the municipality is colored based on its individual BNA score (see Fig. 4.50). The central areas of the city, as indicated by the concentration of greener hexagons,

seem to have higher BNA scores. This suggests better cycling conditions, likely due to a denser network of bike paths, lower traffic stress, or closer proximity to amenities and destinations. In contrast, the periphery of the city, as shown by the red or orange hexagons, appears to have lower BNA scores. These areas might be characterized by fewer cycling infrastructures, different geomorphological conditions (e.g. mountains) or they may be affected by higher traffic speeds and volumes, which can deter cycling. The overall score places the city in a moderate range regarding bikeability. An approximation of the distribution of the number of hexagons for the

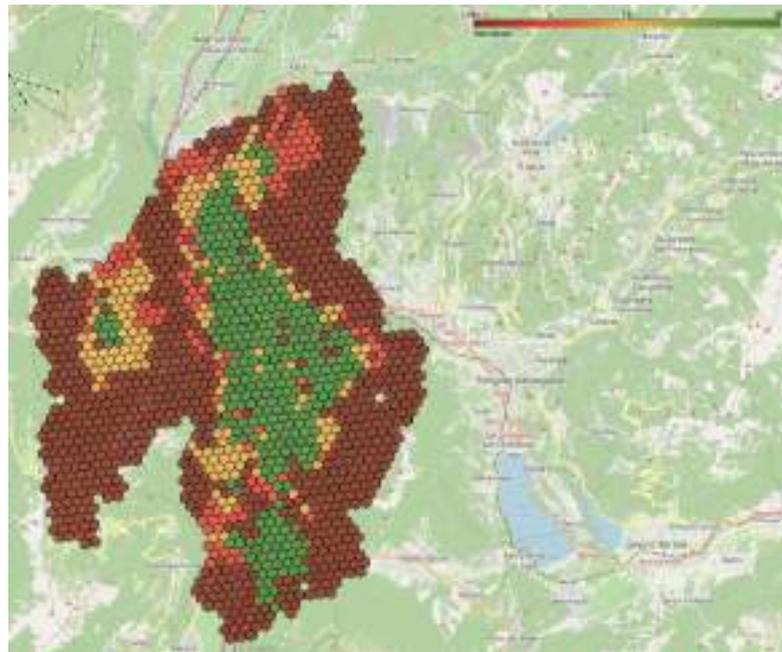


Figure 4.50: H3 hexagon grid of Trento colored according to BNA scores. Green colored hexagons are the ones with higher BNA scores as it can be seen from the legend on the top-right.

rounded total score of each hexagons can be seen here 4.51. Each bar represents the number of

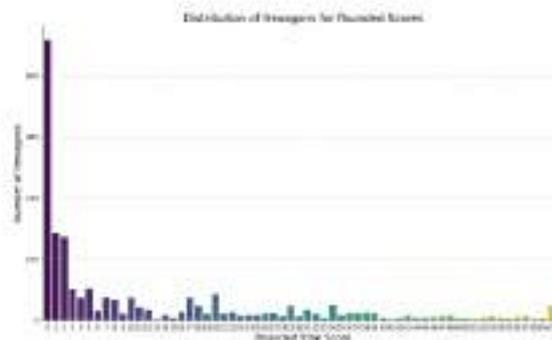


Figure 4.51: Distribution of Hexagons for Rounded BNA Scores

hexagons for a specific rounded total score. The height of the bar corresponds to the count of hexagons with that score. As it can be seen the majority of hexagons has values between 0 and 2, which can be deducted also from the map 4.50. By computing some descriptive statistics we also found that the maximum total score value is 63, the median is 3, the average value is 11.35 and the standard deviation is 16.1. This overall indicates room for improvement, especially in extending the cycling infrastructure to the outskirts and enhancing connectivity between high and low-scoring areas. Regarding Bolzano, the city obtain an overall BNA score of 50.85. Each hexagon of the municipality is colored too based on its individual BNA score (see Fig. 4.53), calculate using also population data (see Fig. 4.52).



Figure 4.52: Population distribution in Bolzano H3 grid, resolution 200m. Red areas are the one more populated (more than 2% of population), followed by yellow one (more than 1% of population) and green one.

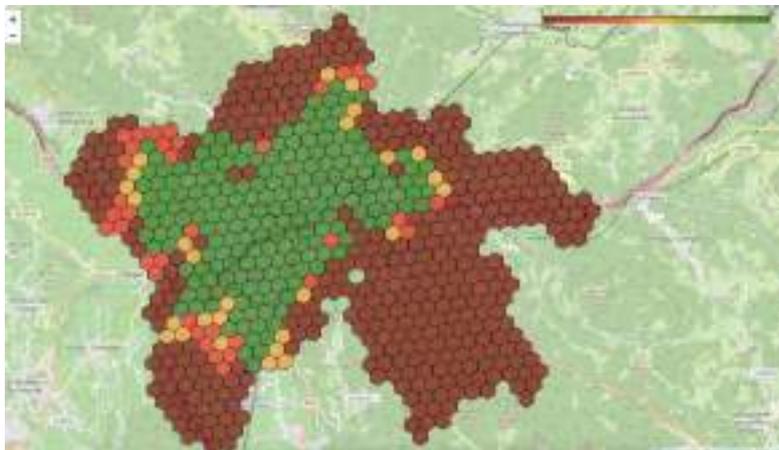


Figure 4.53: H3 hexagon grid of Bolzano colored according to BNA scores. Green colored hexagons are the ones with higher BNA scores as it can be seen from the legend on the top-right.

In the urban core of the city, encompassing neighborhoods like Centro Storico, San Quirino, Europa Novacella, Don Bosco, and portions of Oltrisarco, there is a notable clustering of greener hexagons. This pattern suggests these areas have elevated Bicycle Network Analysis (BNA) scores, indicative of lower traffic stress and a closer proximity to amenities and destinations.

Areas on the peripheries instead are characterized by fewer cycling infrastructures and different geomorphological conditions (e.g. mountains) so the BNA is way more lower. The overall score places also this city in a moderate range regarding bikeability, with a room for improvement. The distribution of the number of hexagons for the rounded total score of each hexagons can be seen here 4.54. Several areas show values between 0 and 1. Descriptive statistics highlights that the maximum total score value is 58, the median is 1, the average value is 14 and the standard deviation is 18.91.

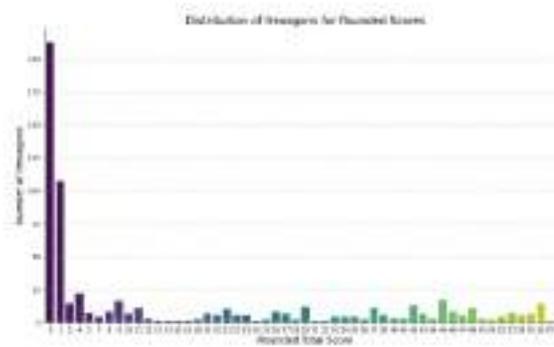


Figure 4.54: Distribution of Hexagons for Rounded BNA Scores in Bolzano

Accident Analysis

In this section, we present an analysis of road accidents that have occurred within the municipal territories of Trento and Bolzano over the past five years, identifying the areas and conditions where these accidents are most prevalent. Regarding Trento, as it can be seen from the Fig. 4.55, the data reveals a notable concentration of accidents primarily in the city center, as depicted in both the scatter plot and kernel density estimation visuals. Particularly, the kernel density estimation plot underscores this pattern by indicating warmer (denser) areas in the central regions. Further analysis uncovers that the majority of traffic accidents occur near streets characterized by a medium level of traffic stress ($LTS = 3$), rather than on those with differing stress levels. Similarly, this trend is observed near intersections. However, an interesting observation is that intersections classified under low or very low traffic stress levels experience a comparable frequency of accidents (see Fig. 4.56). Grouping the roads and intersections by low ($LTS = 1$ or 2) or high levels of stress ($LTS = 3$ or 4) we also note that in general high stress street and intersections show higher number of accidents (see Fig. 4.57). Interestingly, accidents happen mostly on roads with speed limits between 40 and 70 km/h, with an evident peak at 50 km/h and where the number of lanes are equal to two (see Fig. 4.58). Regarding the type of roads involved, the Fig. 4.59 shows that accidents occur in paths of type cycleway, secondary and tertiary, while happen less in primary, residential and pedestrian streets. Statistical tests are used to determine the significance of the differences. In accident frequencies, specifically the Chi-Square test [188] is performed together with pairwise comparison tests. The distribution of lanes and max speed are showed in Fig. 4.60. The Shapiro-Wilk test is conducted to assess the normality of the distribution of lanes and maxspeed. The results indicated that the number of lanes is not normally distributed, $W = 0.51$, $p < .001$. Similarly, the distribution of maxspeed is also not normally distributed, $W = 0.61$, $p < .001$. Then, a Spearman rank-order correlation is conducted to assess the relationship between the number of lanes and the maximum speed

limit in areas where accidents occurred. There is a low positive correlation between the two variables, which was not statistically significant, $r_s(1) = .015$, $p = .228$. A Spearman rank-order correlation analysis is conducted to evaluate the relationship between the number of lanes and the maximum speed limits for different road types within the vicinities of accident sites. Results indicated a small positive correlation for *unclassified* roads, which is not statistically significant, $r_s = .035$, $p = .177$. For *secondary* roads, a moderate negative correlation is found, $r_s = -.290$, $p < .001$. A correlation analysis for *cycleways* is not feasible due to insufficient data variability. This is attributed to the recording of all lanes and maximum speed values in cycleways as NaN. Similarly, no significant correlation is observed for *residential* roads, $r_s = -.013$, $p = .683$; *tertiary* roads, $r_s = -.001$, $p = .963$; and *primary* roads, $r_s = -.067$, $p = .267$. A strong positive correlation is present for *motorway links*, $r_s = .603$, $p = .022$. The analysis for *pedestrian* roads also yields a moderate positive correlation, $r_s = .398$, $p < .001$. Conversely, no correlation is identified for *tracks*, $r_s = 0.0$, $p = 1.0$, indicating no association between lanes and maximum speed. Road types such as *primary link*, *steps*, *tertiary link*, and *living street* does not have enough data to compute a correlation. A Goodness-of-Fit Chi-Square test is conducted to determine whether observed accident counts significantly differed across various levels of traffic stress (LTS) in streets. The Chi-Square value is found to be 2649.31, indicating a significant difference in accident counts across the different LTS categories, $\chi^2(3) = 2649.31$, $p < .001$. Subsequent pairwise comparisons revealed significant differences between all pairs of LTS categories ($p < .001$). These findings suggest that the frequency of accidents is not uniformly distributed across the different levels of traffic stress, with some levels showing higher or lower accident counts than expected under the assumption of equal frequencies. A Goodness-of-Fit Chi-Square test is also performed to examine whether observed accident counts at intersections differed significantly across various levels of traffic stress (LTS). The test reveals a Chi-Square value of 644.53, which suggests a significant difference in accident counts among the LTS categories, $\chi^2(3) = 644.53$, $p < .001$. Further pairwise comparisons indicate significant disparities between specific levels of traffic stress. There is a significant difference in accident counts between LTS 1.0 and LTS 3.0, $p < .001$; between LTS 1.0 and LTS 4.0, $p < .001$; between LTS 2.0 and LTS 3.0, $p < .001$; between LTS 2.0 and LTS 4.0, $p < .001$; and between LTS 3.0 and LTS 4.0, $p < .001$. These results suggest a non-uniform distribution of accidents at intersections related to the varying levels of traffic stress. A Goodness-of-Fit Chi-Square test is conducted to evaluate the distribution of accident counts across different stress levels (low and high). The Chi-Square statistic is significantly high, $\chi^2(1) = 870.12$, indicating that the observed frequencies of accidents are not equally distributed across high and low stress levels, $p < .001$. Subsequent pairwise comparisons reveal a significant discrepancy in accident counts between high and low stress levels, with a Chi-Square statistic indicating a highly significant difference, $p < .001$. These findings demonstrate a distinct variation in accident occurrences corresponding to stress level, with significantly more accidents occurring in one category over the other.

As illustrated in Fig. 4.61, the data pertaining to Bolzano indicates a significant aggregation of traffic incidents, predominantly in the Centro Storico region and its adjacent areas. This pattern is evident in both displayed graphs. Furthermore, it is observed that a substantial proportion of these accidents tend to occur in proximity to streets and intersections characterized by a Level of Traffic Stress (LTS) of 3. This trend is followed by occurrences in areas with streets and intersections of lower LTS values, as further detailed in Fig. 4.62. It is also observed that streets and intersections with higher stress levels tend to register a greater number of accidents compared to those with lower stress levels, though the discrepancy is not markedly pronounced. (see Fig. 4.63). Additionally, the majority of accidents occur on roads where the speed limits range from 40 to 60 km/h, with a noticeable peak at 50 km/h. This pattern is akin to the situation in Trento, particularly on roads with two lanes (see Fig. 4.64 and Fig. 4.66) and more in paths of type

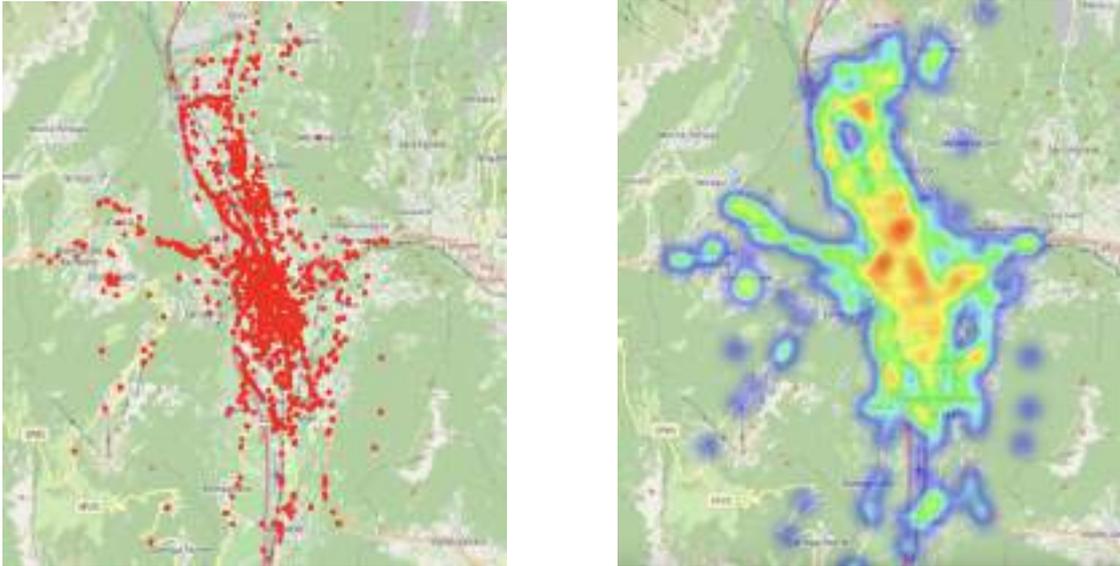


Figure 4.55: Left: Overview of Traffic Accidents in the Trento Municipality (2018-2023). Right: Heatmap of Accident Hotspots in Trento

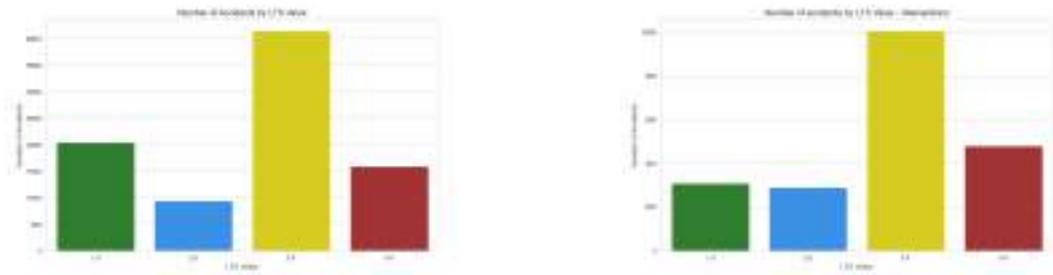


Figure 4.56: Left: Number of Accidents in streets for Level of Traffic Stress in Trento. Right: Number of Accidents in intersections for Level of Traffic Stress in Trento



Figure 4.57: Left: Number of Accidents in streets for Type of Stress in Trento. Right: Number of Accidents in intersections for Type of Stress in Trento

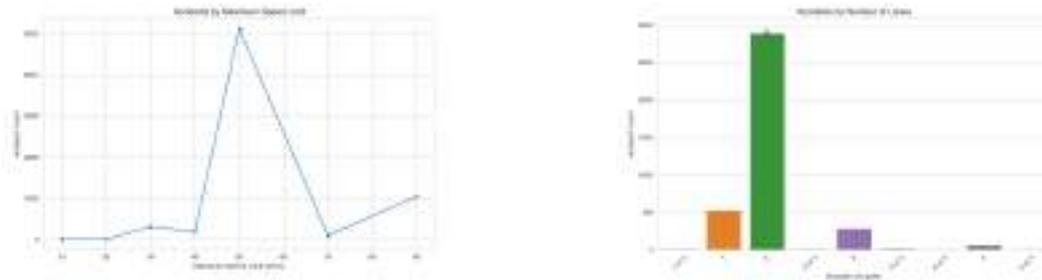


Figure 4.58: Left: Analysis of Traffic Accidents by Speed Limit in Trento Streets. Right: Trento Street Accident Frequency related with the Number of Lanes.

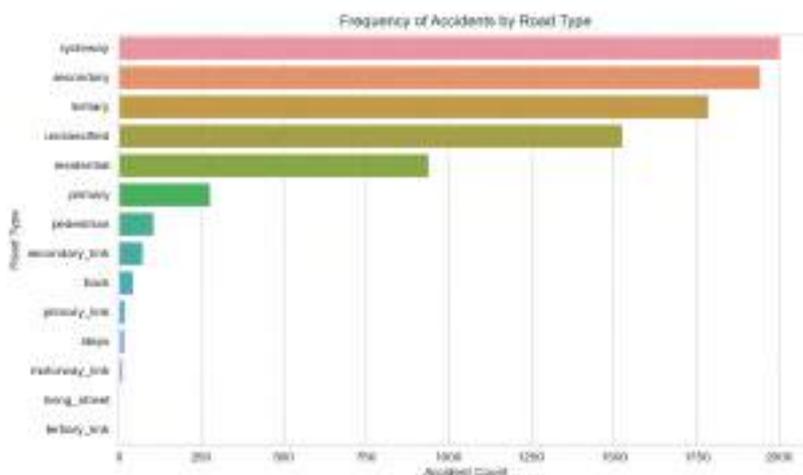


Figure 4.59: Frequency of Accidents by Road Type in Trento

cycleway, residential, secondary and primary (see Fig. 4.65). Results of the Shapiro-Wilk test on the distribution of lanes and maxspeed ($W = 0.74$, $p < .001$) in accidents data indicates that the data for the number of lanes is not normally distributed. The test statistic being significantly less than 1 indicates a deviation from normality. Also the distribution of maxspeed is not normally distributed ($W = 0.71$, $p < .001$). The Spearman rank-order correlation shows that there is a low negative correlation between the two variables, which is statistically significant, $rs(1) = -0.252$, $p < .001$. Then, the Spearman rank-order correlation analysis to evaluate the relationship between the number of lanes and the maximum speed limits for different road types in accidents data shows that there is a moderate negative correlation on primary roads, indicating that as the number of lanes increases, the maximum speed limit tends to decrease ($rs = -0.5919$, $p < .001$). A weak positive correlation is observed for primary_link roads ($rs = 0.131$, $p < .01$). This suggests a slight increase in maximum speed limit with more lanes. A very weak negative correlation exists for residential roads ($rs = -0.05$, $p < .001$). This implies a minor trend where more lanes might correspond to slightly lower speed limits. For secondary roads a moderate negative correlation is found, indicating that more lanes often correlate with lower speed limits

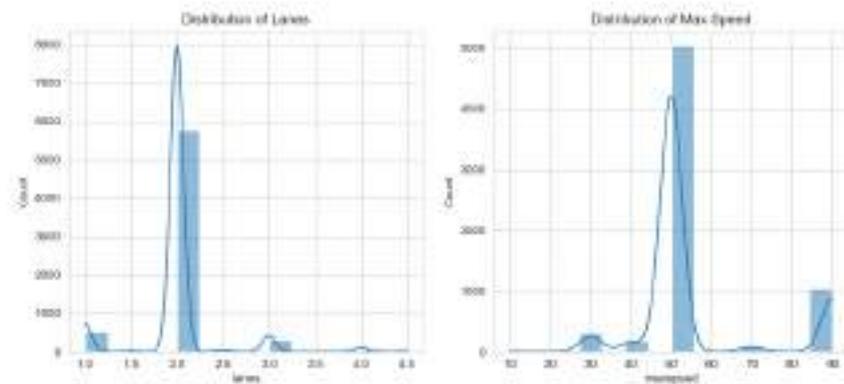


Figure 4.60: Distribution of accidents for different lanes and max speed values in Trento network

($r_s = -0.4342$, $p < .001$). Weak positive correlations are found for unclassified roads ($r_s = 0.152$, $p < .001$) and tertiary roads ($r_s = -0.1365$, $p < .001$). The Chi-Square value is found to be 7116.95, indicating a significant difference in accident counts across the different LTS categories, $\chi^2(3) = 7116.95$, $p < .001$. Similarly to Trento's case. Pairwise comparisons revealed significant ($p < .001$) differences between all pairs of LTS categories. Regarding intersections similar results are found across various levels of traffic stress (LTS). The test reveals a Chi-Square value of 1043.24, which suggests a significant difference in accident counts among the LTS categories, $\chi^2(3) = 1043.24$, $p < .001$. Considering the distribution of accident counts across varying stress levels (low and high), the statistical significance of the Chi-Square test is notable, with $\chi^2(1) = 125.85$. While this significance is less pronounced than in the case of Trento, it still indicates a non-uniform distribution of accidents across high and low stress levels, as evidenced by a p-value less than .001. This outcome highlights a marked disparity in accident frequencies in relation to stress level, clearly showing a predominance of accidents in one category compared to the other.

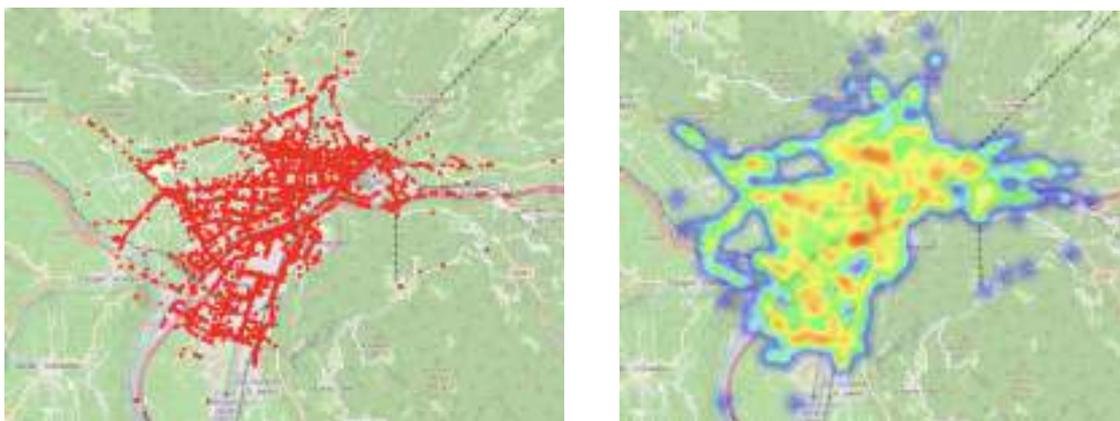


Figure 4.61: Left: Overview of Traffic Accidents in the Bolzano Municipality (2018-2023). Right: Heatmap of Accident Hotspots in Bolzano

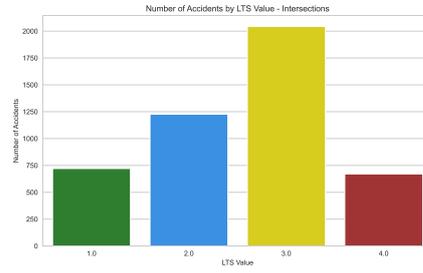
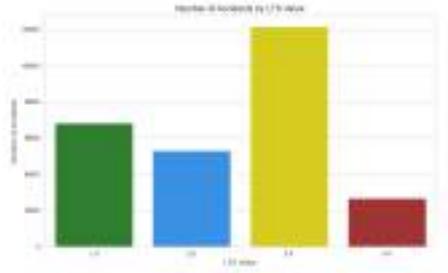


Figure 4.62: Left: Number of Accidents in streets for Level of Traffic Stress in Bolzano. Right: Number of Accidents in intersections for Level of Traffic Stress in Bolzano

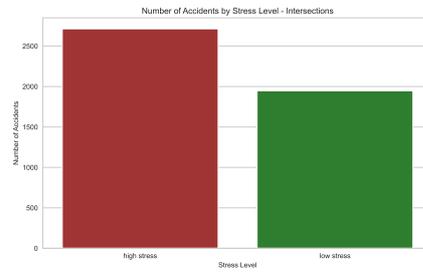
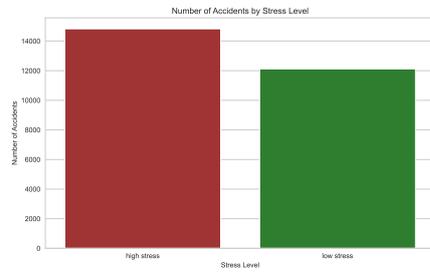


Figure 4.63: Left: Number of Accidents in streets for Type of Stress in Bolzano. Right: Number of Accidents in intersections for Type of Stress in Bolzano

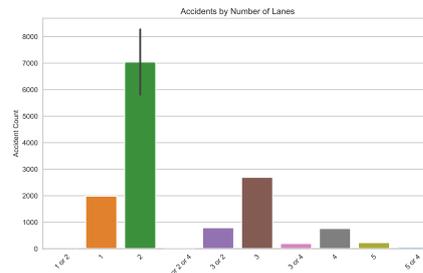
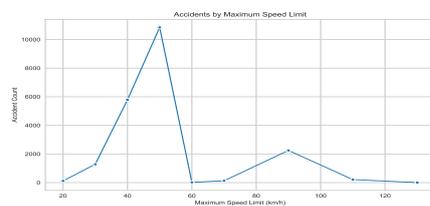


Figure 4.64: Left: Analysis of Traffic Accidents by Speed Limit in Bolzano Streets. Right: Bolzano Street Accident Frequency related with the Number of Lanes.

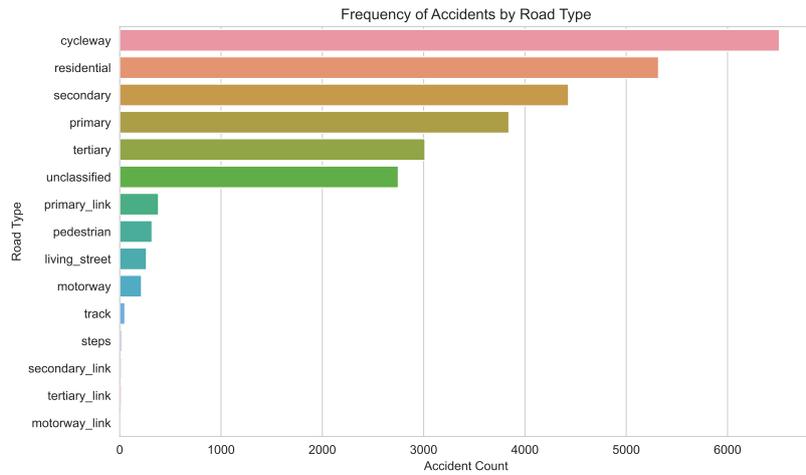


Figure 4.65: Frequency of Accidents by Road Type in Bolzano

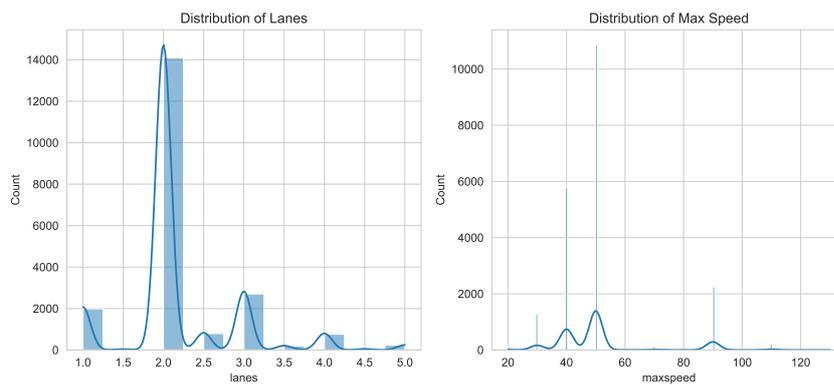


Figure 4.66: Distribution of accidents for different lanes and max speed values in Bolzano network

4.1.5 Final Analysis

This section presents the synthesized findings from the comprehensive analyses conducted before with the goal to identify potential correlations and patterns in the results for both cities.

Spatial Relationship Analysis

Regarding Trento, the Moran’s I analysis revealed significant spatial autocorrelation in specific segments, suggesting that traffic accidents are not randomly distributed but rather concentrated around certain high-stress gaps. This pattern underscores the hypothesis that the presence of infrastructure gaps may contribute to higher accident occurrences. Specifically, the Fig.4.67 shows that the ones in the center and the south part of the city are generally of Quadrant 4, which means that they are associated with high risk of accident but low risk of accidents in the

neighbor areas. A mix of low risk and high risk regards the areas close to the Trento-Sardagna cable car. This is a summary table showing for each component the different counts per quadrant, the average Z-score and the proportion of significant p-values (see Tab. 4.5):

Component	CountQuadrant	Ave. Z-Score per Quadrant	Prop. Sign. P-values
0	{3: 170, 4: 62, 1: 44, 2: 19}	{1: 5.44, 2: -5.43, ...}	0.772881
2	{1: 13, 2: 8, 3: 1, 4: 1}	{1: 2.44, 2: -1.01, ...}	0.695652
7	{4: 2, 2: 1, 3: 1}	{2: 0.0, 3: 1.41, 4: -0.32, ...}	0.250000
10	{4: 2, 2: 1}	{2: -1.0, 4: 0.75}	1.000000

Table 4.5: Summary of components by quadrant in Trento



Figure 4.67: Complete High-Stress Gaps with Varied Degrees of Accident Risk Association.

This output provides a detailed analysis of the spatial autocorrelation of accidents along different segments of the road network (referred to as 'high-stress gaps') using Local Moran's I (LISA). The *Count per Quadrant* contains dictionaries where the keys are quadrant numbers and the values are the counts of components falling into each quadrant. For example, in component 0, there are 170 instances in quadrant 3, 62 in quadrant 4, 44 in quadrant 1, and 19 in quadrant 2. Regarding *Average Z-Score per Quadrant*, it consists in dictionaries mapping quadrant numbers to their corresponding average Z-scores. Z-scores indicate how many standard deviations an element is from the mean. For instance, in component 0, quadrant 1 has an average Z-score of approximately 5.386, and quadrant 2 has an average Z-score of approximately -5.395. Then, in the *Proportion of Significant P-values* each value is a proportion representing the fraction of P-values within a component that are below the defined significance threshold (0.05 in this case). A higher proportion indicates more significant P-values. For example, component 0 has a proportion of approximately 0.773, meaning that 77.3% of the P-values in this component are considered significant. To evaluate the spatial autocorrelation across the entire dataset, the Global Moran's I statistic was computed, serving as an overarching measure that complements

the localized insights gained from the Local Indicators of Spatial Association (LISA). The results indicated a Global Moran's I value of 0.221, suggesting a positive spatial autocorrelation, with a statistically significant P-value of 0.001. This signifies a non-random spatial pattern within the dataset, supporting the presence of clustering in the geographic distribution of the variables under study in Trento. It is also useful to report the proportion of significant local Moran's I statistics based on the chosen significance level which is approximately 0.757, namely the 75.7% of the Local Moran's I statistics are significant. This high proportion further supports the presence of spatial autocorrelation in the data, indicating that a considerable number of segments within the network are part of statistically significant clusters where the pattern of accidents is non-random. Then, in the second graph only the previously prioritized and classified gaps are taken into account (see Fig. 4.68)



Figure 4.68: Declustered High-Stress Gaps with Varied Degrees of Accident Risk Association.

According with the results, the gap that includes the Ravina Bridge (Ponte's Street), the traffic circle and Al Desert Street, and the traffic circle between Hubert Jedin Street and Alcide De Gasperi Street is in Quadrant 1 (brown color, High-High), namely the segment has a high number of accidents and is surrounded with high number of accidents. A similar scenario happens also for the other two gaps of South Trento, namely the one between the traffic circle at Verona Avenue/Madonna Bianca Street and the traffic circle between Enrico Fermi Street and Verona Avenue and the one between Padre Eusebio Chini Street, the traffic circle mentioned above and another traffic circle at Enrico Fermi Street, Alcide De Gasperi Street. Another gap associated to this quadrant is the one that starts in the Trento/Sardagna cable car area near the traffic circle, passing along Lungadige Giacomo Leopardi and Michelangelo Buonarroti Street to the intersection with Dosso Dossi Street, The gap located between Via Vittorio Veneto, Via Dei Muredei and Via Giuseppe Giusti in the San Pio X area, and the section that goes back over Via Giuseppe Giusti to a side road near the Celestino Endrici Archbishop's College is associate with Quadrant 3 (orange: Low-high), namely the segment has a low number of accidents but is surrounded by high number of accidents. The same happens for the gap that starts from the intersection of Milano Street and 3 Novembre Street, passes through the Cavallegeri Bridge, Verona Avenue, Gocciadoro Street and the traffic circle with Paolo Orsi Street in the Santa

Chiara Hospital area. Finally, the sixth and seventh gaps pertain to a highly central zone, which corresponds to the area spanning Giuseppe Verdi Street, Antonio Rosmini Street, and Dei Travai Street, extending to Fiera Square are in Quadrant 4 (High-Low), namely the segments have a high number of accidents but is surrounded by segments with low number of accidents.

In Bolzano, Moran's I analysis has uncovered significant spatial autocorrelation in particular city segments, especially around the central area. This is explicitly illustrated in Fig.4.69. Notably, the central zones (Europa Novacella, Centro Storico, Gries, San Quirino) and the southern part of the city (Oltrisarco) exhibit this pattern. These areas predominantly fall into Quadrant 2 or 3, indicating a low risk of accidents in these areas themselves but a high or low risk in neighboring areas. The count of different quadrants, the average Z-score, and the percentage of significant p-values, are shown in Tab. 4.6:

Component	CountQuadrant	Ave. Z-Score per Quadrant	Prop. Sign. P-values
0	{1: 96, 2: 86, 3: 60, 4: 11}	{1: 3.259, 2: -2.057, ...}	0.624506
1	{1: 6, 3: 6, 2: 4}	{1: 2.204, 2: -2.142, ...}	0.875000

Table 4.6: Summary of components by quadrant in Bolzano



Figure 4.69: Complete High-Stress Gaps with Varied Degrees of Accident Risk Association.

To assess spatial autocorrelation throughout the dataset, the Global Moran's I statistic was calculated. The Global Moran's I value of 0.21 indicates a slight positive spatial autocorrelation, statistically significant with a P-value of 0.001. This implies non-random clustering in Bolzano's geographical data distribution. Approximately 64% of the local Moran's I statistics are significant at the chosen level, underscoring the data's spatial autocorrelation. This indicates that many network segments belong to significant clusters with non-random accident patterns. The second graph considers only the previously identified and categorized gaps (see Fig. 4.70). Within Bolzano's traffic network, two significant high-stress gaps have been identified and are highlighted in Fig. 4.68. The first prominent gap extends from the Resia Bridge, along Resia Street, to Malles Street. It falls into Quadrant 3 (orange color, Low-High), indicating that the segment experiences a low number of accidents and is situated amidst areas with a similarly low accident count. The



Figure 4.70: Declustered High-Stress Gaps with Varied Degrees of Accident Risk Association.

second major gap involves the Talvera Bridge, serving as an important connection between the Centro Storico (historical center) and San Quirino areas. As shown, the overall situation in Bolzano appears less critical compared to Trento, both in terms of the number and the nature of the gaps.

Risk Area Prediction

The analysis is aimed at predicting high-risk accident areas within the network. The goal is to categorize regions as high or low risk based on the accident count by considering as features the BNA total_score and the centrality measures (degree, betweenness, closeness). Furthermore, due to the established presence of spatial autocorrelation in accidents, the inclusion of spatial lags of predictors as additional predictors is also implemented. This section outlines the outcomes of the applied models (Random Forest Classifier (rfc), Support Vector Machine (svm) and Logistic Regression (logreg)) and their performance in classifying the risk levels. Regarding model performance, the resampling techniques used to address data imbalance in the training set included SMOTE, ADASYN, BorderlineSMOTE, and SMOTEENN. The best hyperparameters for each model are determined using GridSearchCV with stratified k-fold cross-validation based on the ROC-AUC score. The following parameters are found to yield the best performance for the respective models:

- Logistic Regression (ADASYN, BorderlineSMOTE, SMOTEENN): $C = 0.1$, solver='lbfgs'
- Random Forest Classifier (SMOTEENN): max_features='log2', n_estimators = 50
- Support Vector Machine (SMOTE, ADASYN, BorderlineSMOTE, SMOTEENN) : $C = 10$, kernel='rbf'

The performance of each model was evaluated based on precision, recall, F1-score, accuracy, precision-recall AUC, and ROC-AUC score. The following results were observed for the logistic regression (see Tab.4.7). The Precision-Recall AUC is 0.80 and the ROC-AUC Score is 0.92.

Regarding the Random Forest classifier, the results can be found here (see Tab. 4.8). The Precision-Recall AUC is 0.79 and the ROC-AUC Score is 0.93.

	Precision	Recall	F1-Score	Support
Class 0	0.98	0.61	0.75	256
Class 1	0.28	0.95	0.43	41
Accuracy			0.66	297
Macro Avg	0.63	0.78	0.59	297
Weighted Avg	0.89	0.66	0.71	297

Table 4.7: Classification report for Logistic Regression

	Precision	Recall	F1-Score	Support
Class 0	0.97	0.90	0.93	256
Class 1	0.56	0.78	0.65	41
Accuracy			0.89	297
Macro Avg	0.76	0.84	0.79	297
Weighted Avg	0.91	0.89	0.89	297

Table 4.8: Classification report for Random Forest Classifier

Regarding the Support Vector Machine classifier, the results can be found here (see Tab. 4.9). The Precision-Recall AUC is 0.61 and the ROC-AUC Score is 0.86.

	Precision	Recall	F1-Score	Support
Class 0	0.97	0.68	0.80	256
Class 1	0.31	0.88	0.45	41
Accuracy			0.71	297
Macro Avg	0.64	0.78	0.63	297
Weighted Avg	0.88	0.71	0.75	297

Table 4.9: Classification report for Support Vector Machine

Regarding the Logistic Regression model, it exhibits an high recall (0.95) but low precision (0.28) for the high-risk category (1). This implies that while the model is good at identifying most high-risk areas, it also incorrectly labels many low-risk areas as high-risk (many false positive). Regarding the accuracy, it is 66%, indicating a moderate level of correct predictions. The F1 score at 43% is more balanced for class 1, indicating a trade-off between precision and recall. The high ROC-AUC Score (0.921) suggests good separability, but this might be misleading due to the imbalance in precision and recall. The Precision-Recall AUC is decent (0.800), indicating moderate capability in distinguishing between classes. Generally, we can see that the model is biased towards predicting class 1, causing many false alarms while successfully catching most of the true class 1 cases. Regarding the Random Forest Classifier, it demonstrates better precision

(0.56) for the high-risk category, indicating fewer false positives compared to logistic regression. It also maintained a high recall (0.78). The accuracy is higher at 89%, showing a better overall predictive ability compared to logistic regression. The F1-score for class 1 is balanced, indicating a good trade-off between precision and recall. Similar to logistic regression the ROC-AUC score is 0.926, indicating good model performance. Also the Precision-Recall AUC is similar to Logistic Regression (0.805). Overall, the Random Forest Classifier shows a more balanced performance between the two classes, with fewer false positives and good accuracy. Regarding the Support Vector Machine, it shows low precision (0.31) for high-risk areas but high recall (0.88), similar to logistic regression, indicating a tendency to overpredict high-risk areas (more false positives). The accuracy stands at 71%, which is higher than logistic regression but lower than the random forest classifier. The ROC-AUC Score (0.861) is good but not as high as the other models. Finally, it has the lowest Precision-Recall AUC (0.610) among the models, suggesting significant challenges in balancing precision and recall. Overall, Support Vector Machine, like Logistic Regression, leans towards predicting class 1, resulting in many false positives but capturing most of the true class 1 cases. Overall, all models struggle with precision, likely due to the imbalanced nature of the dataset (fewer high-risk areas compared to low-risk areas by default (85% percentile threshold)). Random Forest Classifier emerges as the most balanced model, with higher precision and recall and the highest accuracy. The high ROC-AUC scores across models indicate general good discriminatory ability between high-risk and low-risk areas. However, precision-recall trade-offs value reflects the trade-off between precision and recall. While not exceptionally high for Random Forest Classifier, it does indicate that the model has a decent balance between these metrics. In our case, since avoiding missing any high-risk areas is crucial but at the same time is important to reduce false alarms, the model with higher recall and providing less false alarms is the best. For these reasons, the random forest model is selected as the best choice between these three models. Subsequently, the model is retrained utilizing the entirety of the dataset, yielding the following results (see Tab. 4.10). The Precision-Recall AUC is 0.80 and the ROC-AUC Score is 0.94.

	Precision	Recall	F1-Score	Support
Class 0	0.95	0.93	0.94	256
Class 1	0.62	0.71	0.66	41
Accuracy			0.90	297
Macro Avg	0.78	0.82	0.80	297
Weighted Avg	0.91	0.90	0.90	297

Table 4.10: Classification report for Random Forest Classifier - Full Dataset

The ROC-AUC graph is visible here (see Fig. 4.71).

The model identified 189 areas as having a high risk of accidents and 1,295 areas as having a low risk of accidents. Then, SHAP (SHapley Additive exPlanations) is used to interpret the Random Forest model and explain the impact of each feature on the model (see Fig. 4.72).

The SHAP figure illustrates the following information:

- Feature Importance: the features are listed in the order of importance from top to bottom. The feature at the top (closeness centrality mean spatial lag) has the highest impact on model output, followed by the betweenness centrality mean, the BNA total score spatial lag and the others predictors. The feature at the bottom (degree centrality mean) has the least

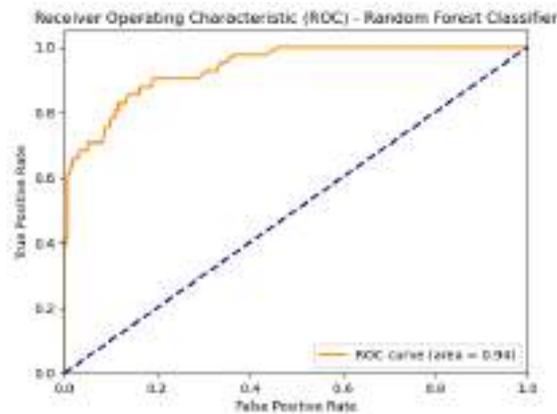


Figure 4.71: ROC-AUC curve - Random Forest Classifier (Trento)

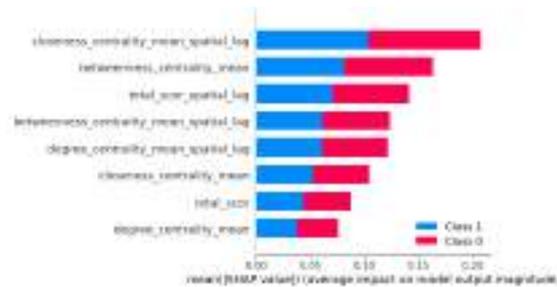


Figure 4.72: SHAP output for RFC model (Trento)

impact among the listed features.

- **Impact Direction:** each feature has two colors representing the two classes (Class 0 and Class 1). The color signifies the direction of the feature's impact, which are red for Class 1 (high risk areas), where higher values of the feature push the model's prediction towards class 1 and blue for Class 0 where higher values of the feature push the model's prediction towards class 0.
- **Impact Magnitude:** the horizontal location shows the impact magnitude (the mean absolute SHAP value for a feature). A feature's impact is larger if its average SHAP value is further from zero. For instance, a high value of 'closeness_centrality_mean_spatial_lag' contributes significantly to shifting the prediction towards Class 1.
- **SHAP Value:** on the x-axis, the SHAP value represents the mean absolute impact on the model output. This is an average measure across all the samples. It's not about how high or low the feature value is; rather, it's about how much the prediction shifts when the feature is present.

The Closeness_Centrality_Mean_Spatial_Lag exerts the most substantial influence on the model's predictions, with higher values predominantly favoring Class 1. This is closely followed by Betweenness_Centrality_Mean, which displays a diverse range of impacts; both its lower and higher values significantly affect predictions across both classes. Subsequently we have the spatial lags

related to BNA Total Score, Betweenness Centrality and the Closeness Centrality Mean. Lastly, the BNA Total_Score and the Degree_Centrality_Mean hold the least impact, characterized by a tighter spread of SHAP values and a subtler effect on the model's predictive outcomes. Finally, a map is created for displaying areas of different level of risk for accidents and can be seen here (Fig. 4.73).

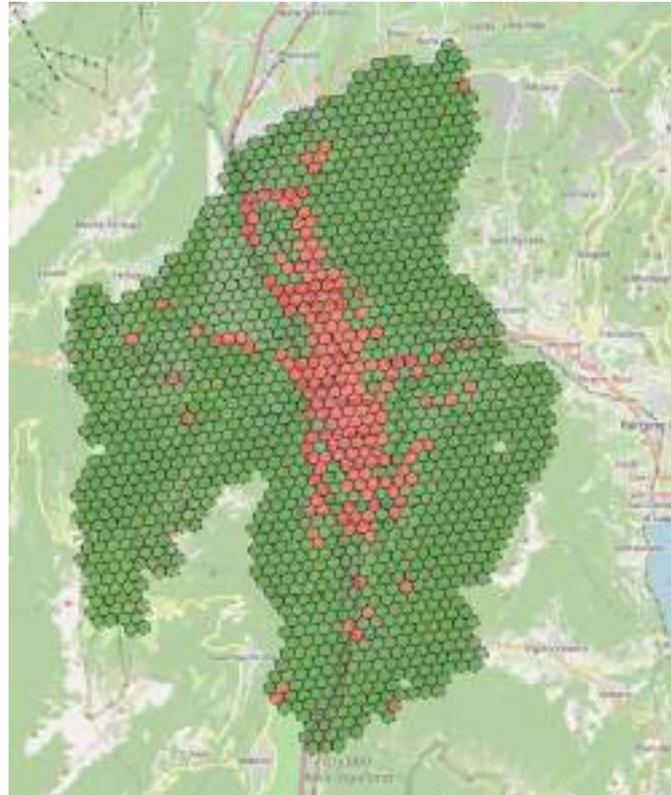


Figure 4.73: Zones in Trento categorized by high or low accident risk.

In examining the spatial distribution of traffic accident risk as presented in Figure 4.73, it is observed that areas of high risk are predominantly centralized within the urban core of Trento. This distribution aligns longitudinally with the primary arterial routes extending from the southern to the northern sectors of the city. Additionally, isolated clusters of high-risk zones are discernible adjacent to peripheral urban concentrations and along major transportation corridors. Regarding Bolzano, the Random Forest Classifier model yields the following results (see Tab. 4.11). The Precision-Recall AUC is 0.93 and the ROC-AUC Score is 0.96.

The ROC-AUC graph is visible here (see Fig. 4.74). The model identified 184 areas as having a high risk of accidents and 439 areas as having a low risk of accidents. Then, SHAP (SHapley Additive exPlanations) is used to interpret the Random Forest model and explain the impact of each feature on the model (see Fig. 4.75).

Based on the SHAP analysis, also in this case among the most influential features we find the Betweenness_centrality_mean, succeeded by the BNA_total_score. Then we have the degree_centrality_mean_spatial_lag and the other predictors. Also here, the degree_centrality_mean is the one that provides lower impact together with BNA total score and Closeness_centrality_mean.

	Precision	Recall	F1-Score	Support
Class 0	0.96	0.93	0.95	451
Class 1	0.84	0.90	0.87	172
Accuracy			0.92	623
Macro Avg	0.90	0.91	0.91	623
Weighted Avg	0.93	0.92	0.92	623

Table 4.11: Classification report for Random Forest Classifier - Full Dataset

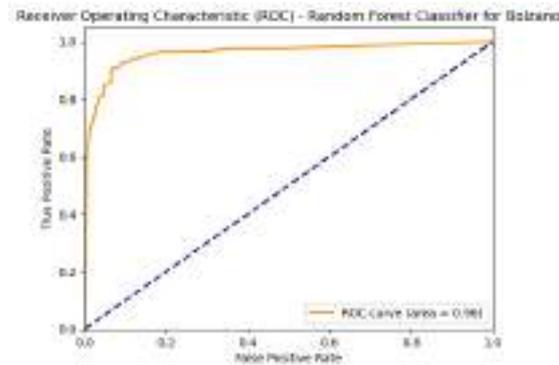


Figure 4.74: ROC-AUC curve - Random Forest Classifier (Bolzano)

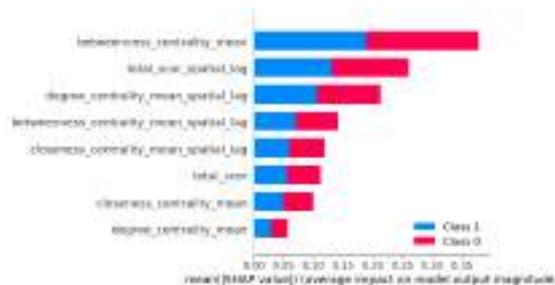


Figure 4.75: SHAP output for RFC model (Bolzano)

The magnitude of contribution from the features is substantially larger than that of the other scenario, significantly impacting the model's output. A map illustrating areas with varying accident risk levels has been developed and is available for viewing (Fig. 4.76). Referring to Fig. 4.76, it is evident that high-risk areas are primarily concentrated in the urban core of Bolzano, with their prevalence diminishing progressively as one moves away from this central region. This spatial distribution correlates closely with the urbanized sectors of the municipality. Additionally, a handful of isolated high-risk clusters can be observed near the hospital area and along the major primary roads.



Figure 4.76: Zones in Bolzano categorized by high or low accident risk.

4.2 Discussion

In initiating the discussion of this paper, it is pivotal to revisit the core research questions that have guided our exploration into the intricacies of urban cycling networks. The primary aim of this research is to delve into the dynamics of urban cycling, with an emphasis on traffic stress, infrastructures, accessibility and their intersection with urban accidents. Through the following questions, the research aims to dissect the complex nature of urban cycling networks, illuminating how various factors intertwine and influence the decisions and safety perception of cyclists. This investigation is steered by several key inquiries. Firstly, we seek to identify and categorize areas of varying traffic stress within urban landscapes of diverse geomorphological features. This involves determining which streets or zones are subject to high traffic stress and which exhibit lower levels of stress, and whether there are discernible clusters of high traffic stress within the city. Secondly, the research focuses on evaluating accessibility and proximity to essential destinations in the urban landscape. This encompasses identifying streets and areas with enhanced accessibility and those that are conveniently located near common destinations, such as services and amenities, and assessing the overall state of accessibility across the area. Thirdly, we examine whether the urban cycling networks consist of subgraphs characterized by distinct levels of traffic stress. This includes identifying the number of high-stress gaps in areas of low stress and determining the significance of these gaps. The fourth question investigates the spatial correlation between the locations of accidents and high traffic stress gaps within the urban network. This seeks to understand if there is a significant geographical relationship between accident sites and areas of high traffic stress. The insights garnered from these inquiries are anticipated to significantly contribute to the creation of more cyclist-friendly and safer urban environments. The last query revolves around the predictive analysis of areas at varying risks of accidents. This involves integrating quantitative metrics, like centrality measures, with qualitative evaluations such as accessibility scores and traffic stress levels, to foresee potential high and low-risk zones for accidents. Given the presence of significant spatial autocorrelation in road accidents, spatial lags of the predictors have been incorporated into the predictive analysis. Regarding the first research

question, our analysis revealed that in both cities, Level 3 Traffic Stress (LTS 3) is the most prevalent category on streets and intersections and this is in line to what found by Moran et al. [168]. This is followed by LTS 1 or LTS 2, with LTS 4 being the least common. The dominance of Level 3 Traffic Stress in both cities suggests that a significant portion of streets and intersections have intermediate levels of traffic stress. This implies that while these areas may not be the most challenging for cyclists, they still present notable barriers to safe and comfortable cycling, especially for less experienced riders. We observed some consistent patterns in the distribution of these traffic stress levels. In both cities, LTS 4 areas are typically found in two distinct types of locales: one characterized by elevated, sloping terrain with less dense road networks, and the other associated with industrial zones or areas known for traffic congestion. This is supported by the clustering results and high stress connected components analysis which corroborate the observation that areas of high traffic stress are frequently located near busy urban zones, industrial regions and areas with significant slope. These findings suggest a correlation between high traffic stress and specific urban characteristics, including dense traffic, industrial activities, and challenging topography. Low stress streets (LTS 1 and LTS 2) are more frequently situated closer to city centers and residential areas. In contrast, LTS 3 areas are dispersed throughout the cities, often marking busier streets that lack specialized cycling infrastructure. In fact, according to Moran et al. [168] this type of LTS streets is the one that needs to be transformed to provide the greatest benefit to the overall bicycle network. The trend of lower stress levels (LTS 1 and LTS 2) being more prevalent in city centers and residential zones suggests a trend towards environments more conducive to cycling in these areas. This pattern is supported by numerous examples, such as Toronto, Paris, London, Barcelona, New York, among others [124, 187, 205]. Locations with fewer lanes, restricted traffic areas, and reduced traffic speed seem to be more attractive for cyclists so they correspond to lower levels of traffic stress [85, 124, 160]. A US regional screening tool for bicycle LTS and connectivity analysis Delaware Valley Regional Planning Commission [71] reinforces this understanding. It suggests that low-stress areas (LTS 1 and LTS 2) often form *'islands'* in city centers or residential areas. The analysis implies that these low-stress areas are comfortable for most cyclists, but their connectivity to other parts of the city might be limited by higher stress roads, such as the LTS 3 and LTS 4 [70, 168] and this is reflected in our two cases. On the other hand, the lower presence of high stress levels (LTS 4) suggests that challenging conditions for cyclists are less common, but where they do exist, they pose significant problems in terms of safety and accessibility, as stated also by Lowry, Furth, and Hadden-Loh [150]'s work. Consistently with this, Mekuria, Appleyard, and Nixon [159] and Kent and Karner [134] found that LTS in industrial areas usually go from 3 to 4. Moreover, in Abad and Van der Meer [1] work it can be seen that high-stress paths (LTS 3 and 4) found are located in low accessibility, impervious, and sloping areas. The fact that LTS 3 areas are spread throughout the cities indicates a generalized need for improvements in cycling infrastructure beyond the city centers, as highlighted by [71, 168]. This dispersion suggests that while central areas may be becoming more cyclist-friendly, other parts of the city are lagging behind. Regarding the second question, destination accessibility analysis in both network provide several interesting findings. Trento has an overall BNA score of 46.58, placing it in a moderate range for bikeability, where central areas of Trento have higher BNA scores compared to the others. These areas have better cycling conditions due to a denser network of bike paths, lower traffic stress streets, or closer proximity to common destinations. On the other hand, the outskirts of the city have lower BNA scores due to the presence of less cycling infrastructure, challenging geomorphological conditions, or higher traffic speeds and traffic volumes. Statistical analysis show a significant variation in bikeability across the city probably due to the fact that the municipality includes several areas less prone to bicycle use, such as mountains. Conversely, Bolzano's overall BNA score of 50.85 suggests moderate bikeability. Similarly, the

urban core exhibits higher BNA scores, reflecting better cycling conditions, whereas the peripheries, with limited cycling infrastructure and challenging geomorphological features, score lower on the BNA scale. This situation also displays a broad spectrum of bikeability levels across different areas. Overall, both cities display a similar trend of higher bikeability in central areas and challenges in the peripheries, indicating a need for targeted urban planning and infrastructure development to improve overall cycling conditions. While it is true that excluding non-urban, mountainous areas lacking infrastructure from the analysis could potentially yield higher Bicycle Network Analysis (BNA) scores, it is crucial to consider these scores in the broader context of what has been achieved in other cities using similar methodologies. For instance, the study by Abad and Van der Meer [1] on Lisbon, a city known for its slopes, resulted in a considerably lower BNA score of approximately 8.6 out of 100. In contrast, Italian or European cities without such topographical variations demonstrate markedly different values. According to the PeopleForBikes [191] report, Milan scored 69/100, Verona 66/100, and Bologna 73/100. However, it's important to note that in these Italian cases, the impact of slopes on traffic stress levels and subsequently on BNA scores was not factored into the analysis. In addition, areas with slopes are rather limited if not absent in these municipalities. This discrepancy highlights the significance of considering geographical variations and specific urban characteristics when evaluating and comparing BNA scores across different cities. Regarding the third inquiry, our new LTS-adapted IPDC procedure, derived from the work of Handshake Cycling [109], Lowry, Furth, and Hadden-Loh [150], PeopleForBikes [189], Vierø, Vybornova, and Szell [235], and Vybornova et al. [237] and partially from these other works [40, 41, 168, 244]. These indicate the existence of numerous isolated low-stress segments, interspersed with high-stress connections, echoing the observations made by Moran et al. [168] and aligning with the conclusions of Furth [89], which emphasize the necessity of low-stress connectivity in a bicycle network. Moreover, a disparity in the quantity of high-stress traffic gaps identified between the two cities' networks is found. For Trento, a total of eight high-stress traffic gaps have been prioritized. These are predominantly concentrated around the urban core and, by direct knowledge, exhibit peak congestion during specific times of the day. In contrast, Bolzano exhibits three distinct high-stress gaps. These not only encompass areas encircling the city's center but also extend to regions linked with industrial and service sectors and this is in line with what Lowry, Furth, and Hadden-Loh [150] found for Seattle case. Notably, in both cities, the identified high-stress zones frequently incorporate infrastructural elements such as roundabouts and bridges. As similarly found by the work of Vybornova et al. [237] for Copenhagen, the majority of gaps found are related to the type 'street' and 'bridge' and, as suggested by the authors, to enhance the methodology, it should be helpful to introduce additional subcategories within the street gap classification such as road conditions, speed limits or empirical traffic volume data. This could assist in determining the viability of these routes for safe bicycling, despite the absence of dedicated bike infrastructure. This approach aligns with the recommendations put forth by de Groot's work [102]. Then, a comparison of our lists of prioritized gaps is done with the current cities' Biciplans [55, 58]. In Trento, there is a varied picture concerning the overlap with Biciplan projects. Specifically, the gap encompassing the Ravina Bridge (Ponte's Street), the traffic circle at Al Desert Street, and the junction of Hubert Jedin Street and Alcide De Gasperi Street currently has bicycle paths only near the bridge, with the remainder slated for future development. The second gap, located between the traffic circle at Verona Avenue/Madonna Bianca Street and the junction between Enrico Fermi Street and Verona Avenue, and the third gap, situated close by and spanning from Padre Eusebio Chini Street to the traffic circle at Enrico Fermi Street and Alcide De Gasperi Street, already feature bicycle infrastructure as indicated on the map. However, differently from the original work of Vybornova et al. [237], our analysis focuses on identifying high traffic stress gaps within low-stress zones. Therefore, merely the presence of bicycle infrastructures, without

additional context, does not adequately account for the lack of stress in these areas. In my direct experience they turn out to be heavily trafficked areas with sometimes broken and discontinuous bicycle infrastructure. Regarding the fourth gap, located between Via Vittorio Veneto, Via Dei Muredei and Via Giuseppe Giusti in the San Pio X area, and the section that goes back over Via Giuseppe Giusti to a side road near the Celestino Endrici Archbishop's College, we assist to a mixed scenario with already present infrastructure and planned one. The same happens for the fifth gap, namely the one that starts from the intersection of Milano Street and 3 Novembre Street, passes through the Cavallegeri Bridge, Verona Avenue, Gocciadoro Street and the traffic circle with Paolo Orsi Street in the Santa Chiara Hospital area. According to the Biciplan [58], in this case we have a first part with bicycle infrastructures (close to the Cavallegeri Bridge), a part with no infrastructure or planned one, and a second part (the one close to the hospital) with a planned infrastructure. Again from firsthand experience, the bridge portion of the already existing infrastructure involves a crosswalk that is not easy and where, above all, many cars and buses pass by. The gaps identified in the areas includes Giuseppe Verdi Street, Antonio Rosmini Street, and Dei Travaì Street extending to Fiera Square, presents a mix of both existing and yet-to-be-developed bicycle facilities. In contrast, the area beginning near the Trento/Sardagna cable car station and extending along Lungadige Giacomo Leopardi and Michelangelo Buonarroti Street up to the intersection with Dosso Dossi Street, displays a different scenario. The initial segment of this route is not highlighted in the Biciplan, indicating a lack of existing bicycle infrastructure, whereas the latter part already possesses such facilities. The comparison regarding Bolzano's cycling infrastructure reveals a mix of existing and proposed bike paths, particularly around the area stretching from Libertà Street to the Hospital and Merano Street [55]. This sector shows a blend of already existing bike ways alongside planned ones. This suggests that the area is already under consideration for further development, highlighting the recognized need for enhanced cycling infrastructure in these parts. Regarding the two other gaps, these are not specifically highlighted by the Biciplan and this can be explained by the fact that there are alternatives in the project map in the vicinity of these busy routes that would aim to redirect bicycle users to dedicated routes or by the fact that these routes are not considered prioritized by the document. The concluding observation is that these findings are indicative of the variances in urban planning strategies and the availability of bicycle facilities between the two cities, and how these factors influence the extent of low-stress connectivity and related missing links. In examining the spatial correlation between accident locations and high traffic stress gaps in the urban network, Moran's I analysis reveals significant spatial autocorrelation in specific areas of Trento. According to the results, accidents are more concentrated around particular high-stress gaps, as opposed to being evenly distributed across the network. The central and southern parts of the city, identified as Quadrant 4, show a high risk of accidents with lower risk in neighboring areas. Conversely, certain gaps near the Trento-Sardagna cable car are a mix of low and high risk. The global Moran's I value indicates a low positive spatial autocorrelation, signifying non-random accident patterns and clustering. Specific gaps, like those around Ravina Bridge and South Trento, are highlighted for their high accident frequencies. Other areas, such as the San Pio X area and near Santa Chiara Hospital, show contrasting patterns of low accident numbers surrounded by high-risk zones. Regarding Bolzano, a low positive spatial autocorrelation is found too. Central zones (Europa Novacella, Centro Storico, Gries, San Quirino) and the southern part of the city (Oltrisarco) exhibit patterns similar to Trento. Specifically, these areas generally present low risk of accidents in the segments and high/low risk of accidents in neighboring areas. Interestingly, only some segments in the southern part of the city (industrial) and eastern one present Quadrant 1 streets, meaning high number of accidents. Finally, the filtered list of gaps previously found with the adapted IPDC procedure for Bolzano, present low risk of accidents in the segments themselves and high risk of accidents for neighboring areas.

Overall, this analysis is crucial for understanding the geographic relationship between accident sites and high traffic stress areas. Since sustainable mobility and bicycle plans (Biciplan) aim to reduce accidents, identifying these gaps where accidents frequently occur is invaluable [55, 58]. To evaluate the variation in accident risk levels across different areas, a comprehensive method that merges quantitative and qualitative assessments is utilized. This approach includes the integration of centrality measures, accessibility scores, traffic stress levels and, as mentioned above, their relative spatial lags. If we've already discussed traffic stress levels and BNA scores, then regarding centrality measures the urban structures of Trento and Bolzano show both differences and similarities. In Trento, central areas exhibit higher levels of all centrality measures (degree, closeness and betweenness) compared to peripheral regions. Bolzano's pattern is similar but with a more subdued degree and closeness centrality, indicating a well-connected hub. However, in terms of betweenness centrality, Bolzano differs as its high-centrality nodes are more dispersed in the central network than in Trento. Concerning traffic accidents in Trento and Bolzano, a significant concentration is observed in city centers, especially on roads with a medium stress level ($LTS = 3$) around intersections. This finding aligns with what has been discovered by [46, 51]. A significant number of these occur on streets with speed limits between 40 and 70 km/h, with a peak at 50 km/h, and predominantly on two-lane roads. The highest frequencies of accidents are noted on cycleways, secondary and tertiary roads. The decision to predict a dichotomous variable (high/low risk areas) for accident risk assessment, rather than working directly on the number of accidents for areas, is driven by several pragmatic, methodological considerations and benefits in terms of interpretability and alignment with decision-making processes. Primarily, this approach simplifies and clarifies the model's outputs. By categorizing risk into binary 'high' and 'low' classes, the model becomes more user-friendly and interpretable, particularly for stakeholders like urban planners or local authorities. This binary classification aids in decision-making by providing clear-cut categories for resource allocation and intervention strategies, which can be more straightforward than interpreting continuous variables. Moreover, as it can be seen from the accident visualization graphs, many areas might have minimal or no accidents, other might have homogeneous frequency and others might have significantly higher counts. Modeling raw counts could result in models overly influenced by areas with high accident rates, thereby overshadowing the broader risk landscape. Dichotomizing the data helps mitigate the impact of this non-homogeneity by focusing on the risk level rather than the absolute number of accidents. From an operational standpoint, the dichotomous model reflects real-world risk management practices. Decision-makers typically categorize areas into various risk levels to prioritize actions. By modeling risk as 'high' or 'low' the outputs become directly applicable to these real-world scenarios. The decision to set a threshold for high/low risk categorization to 15% (85% percentile) is grounded on the actual distribution of the data, highlighting areas that significantly deviate in terms of risk. The selected Random Forest model identifies for Trento 189 high-risk and 1295 low-risk accident areas. Key results include a precision of 0.95 for Class 0 and 0.62 for Class 1, with an overall accuracy of 0.90. The model achieved a ROC-AUC score of 0.94, indicating high accuracy. The SHAP analysis highlighted the significance of features such as the mean spatial lag of closeness centrality, mean betweenness centrality, and the spatial lag of the BNA total score in predicting different risk classes. According to the map (Fig. 4.73) high-risk areas are mainly in the urban core and along major routes. For Bolzano, the model shows an overall accuracy of 92%. Moreover, The ROC-AUC score is notably high. SHAP analysis indicates that features like betweenness centrality mean and the spatial lag of BNA total score are crucial in influencing the model's predictions. A map detailing the varying risk levels across Bolzano highlights a concentration of high-risk areas in the urban core, aligning with the city's urbanized sectors. Specifically, 184 high risk areas and 439 low risk ones. Overall, this suggests that areas with denser or more central street networks and areas which serve as key connection

points in the network tend to have a higher risk of accidents, likely due to increased traffic and interactions with other modes of transportation in these central urban locations. The results pertaining to intersection density align with this perspective. This is supported by the study of Osama and Sayed [184], which illustrates a positive association between cyclist-related collisions and factors such as the linearity and connectivity of bike networks, underscoring the influence of network structure indicators. Additionally, Prati, Pietrantoni, and Fraboni [196]’s work corroborates the significance of urban streets in the occurrence of accidents. Furthermore, accident data from both cities reveal a predominant occurrence near cycle paths, potentially explaining their association with higher BNA values (see 4.65 and 4.59). Moreover, the prominence of spatial lags of some centrality measures (specifically betweenness and closeness) as key predictors highlights the significance of not only the specific location’s traffic flow and accessibility but also the impact exerted by its neighboring regions. This insight is pivotal for deciphering road accident patterns, as traffic dynamics and associated risks extend beyond solitary locations, influenced considerably by adjacent areas. These spatial lags adeptly encapsulate such interdependencies, rendering them essential in identifying areas with a high risk of accidents, as found by Al-Hasani, Asaduzzaman, and Soliman [113] and Wang et al. [240]. The importance of this finding lies in its practical application: recognizing that high-risk areas are affected by their surroundings can lead to more targeted and effective intervention strategies. For instance, enhancing traffic management in an area characterized by high betweenness centrality could potentially mitigate accident risks in that area and its neighboring zones. Building on the aforementioned results, a key inference for exploratory accident analyses and the identification of high-risk areas is that roads or routes characterized as low-stress do not invariably align with reduced accident rates. This suggests that even though some roads are considered less stressful for travel, such as those with a maximum speed limit of 30-50 km/h, it doesn’t necessarily lead to a decreased probability of accidents. In fact, the highest incidence of accidents is observed in areas where the speed limit is 50 km/h. This goes in line with the idea that the mere presence of a bicycle infrastructure, even if isolated from the mixed-traffic lanes, is not the sole and sufficient condition for safer streets with low perception of danger and stress and thus prompting users to move actively, but that an important factor is the quality of this infrastructure [16, 110, 119, 221, 243]. By pinpointing these high-risk areas, urban planners and policymakers can target interventions more effectively, enhancing safety and achieving the goals of sustainable mobility initiatives. This approach not only aids in making more informed decisions for infrastructure improvements but also contributes significantly to creating safer, more cyclist-friendly urban environments [55, 59, 84, 222, 225].

Chapter 5

Limitations and Future Research

5.1 Limitations and Future Research

This chapter is dedicated to discussing the key limitations of the study while also offering recommendations for future research or practical applications of the quantitative and qualitative methods used to evaluate safe and accessible bikeability. This approach aims to enhance the effectiveness of future studies in this domain by building on the learnings and insights gained from the current research. As mentioned in the introduction, the accuracy and update frequency of OpenStreetMap data may vary or may lack, potentially affecting the robustness and the completeness of the results [115], which is particularly pronounced in less renowned areas [17]. Although the OpenStreetMap community has established tagging guidelines, their strict adherence varies and is not consistently enforced [251]. The quality of the Bike Network Analysis (BNA) score is heavily dependent on the accuracy of OpenStreetMap (OSM) data, so is susceptible to human errors which is a main limitation [191]. Specifically, future iterations could benefit from cross-verifying with local governmental data. Apart from data concerning population distribution in the areas of interest, accident data and digital elevation model data, this study did not incorporate any external datasets related to the road network from OpenStreetMap. Specifically, data concerning road traffic flow, which could provide additional insights into the most frequented areas and thus indicate where there is a greater need to ensure low levels of traffic stress or reduce accident numbers for improved accessibility and safety perception, was not utilized. This includes datasets that could be sourced from local public administrations or private services. The population data utilized in this study was sourced from a service that offers a general approximation rather than precise figures [136]. The calculation method employed involves the integration of the Global Human Settlement Layer (GHSL) with Facebook's High Resolution Settlement Layer (HRSL) population data, where such data is accessible. This approach combines these two datasets to provide a comprehensive overview of population distribution. Therefore, if updated data on the population density of a specific region were available, it would be advisable to utilize this information for a more accurate analysis. In this study, instead of utilizing Italian census blocks or similar geographic units commonly employed in previous BNA or LTS analyses (for example [239]), we have adopted the H3 Geospatial Indexing System for population data [104, 105]. The H3 system comprises uniformly shaped and sized hexagonal cells that create a global grid. This uniformity in the grid structure enhances the efficiency and precision of geospatial data analysis, allowing for more consistent and reliable spatial assessments compared to the irregular and variable sizes of traditional census blocks. Moreover, according to Furth, Mekuria, and Nixon [90], large spatial unit sizes are not suitable in LTS analysis due to the variability in network

stress levels within a single zone. This aspect is also not to be underestimated when calculating the BNA score, as it assumes uniform low-stress within an analysis unit, regardless of its size. This method risks not accounting for the diverse levels of traffic stress present in larger geographical areas, which might compromise the precision of the study's findings. In our study, due to the unavailability of direct data on bicycle traffic flow, we used a different approach for gap analysis based on its original work [237]. We estimated the most frequently used bicycle routes, thereby identifying critical gaps, through a metric called *'link betweenness centrality weighted by gap length'*. This metric operates on the premise that cyclists typically opt for the shortest route between their starting point and destination. The betweenness centrality of a network link reflects its frequency of use in the shortest paths between all node pairs. By factoring in the gap length, we can approximate the total meters cycled in mixed traffic for each gap. This approach assumes one *'cyclist unit'* for each pair, always taking the shortest route. Integrating IoT data, like from sensors or smart traffic systems, could provide more dynamic and accurate insights into traffic patterns and flow. It could be useful also for analyzing cyclist behavior and subjective stress perception [48, 51, 86, 193]. Regarding the adaptation of the LTS process to Italian urban settings, several assumptions were made in order to deal with the fallacies and the lacks of OSM. In cases where speed limit data or number of lanes were unavailable, assumptions were made based on road type and Italian road laws. For road slope data, we utilized raster data from Tinitaly, calculating and assigning slope values to each road with approximately 10-meter accuracy. Furthermore, roads were classified as urban or extra-urban based on their placement within construction quantiles. This approach ensured a contextual and data-driven categorization of road types in the absence of explicit speed limit information but with the possibility of errors to not underestimate. Future studies could be expanded to include more diverse urban settings within Italy considering a vast scenario of geomorphological conditions and include micro-level urban design features. Other data regarding the quality of the infrastructures or surfaces weren't available. As mentioned several times, it was not possible to use all the variables (e.g. intersection delays, weather, natural beauty etc.) desired by the authors for LTS classification [160] due to their unavailability. However, like other researchers, we have also incorporated a crucial variable into our study that significantly affects risk perception and stress levels: the slope of the roads [1]. This addition of external data enhances the depth and relevance of our analysis, considering the impact of road gradients on cycling experience and safety. In the context of calculating the Bike Network Analysis (BNA) score, workplaces for census blocks were not included among the common destinations due to the lack of specific data for the studied cases. For future research, there is potential to refine the destination parameters by incorporating workplace data as done by PeopleForBikes [191] and Abad Crespo [2], which could yield a more comprehensive understanding of major traffic flows and patterns. This addition would significantly enhance the depth and applicability of the BNA score in reflecting real-world cycling dynamics. Moreover, the influence of destination weights given based on Abad Crespo [2] procedure on the BNA score is a critical aspect, as varying weights can significantly alter the resulting score. Consistency in scoring criteria is essential for accurate comparisons across different study areas. Robust validation process should include assessing the appropriateness of destination weights, ensuring they align with the actual significance, as stated by Abad and Van der Meer [1]. Traffic accident data for this study was sourced from local open data platforms. However, these datasets often do not differentiate the types of accidents, making it challenging to filter specifically for incidents involving cyclists. In some cases, exploratory analysis enabled the identification of accidents occurring on or near bicycle paths. This approach, while useful, introduced the potential for increased error in the assessment due to the indirect method of determining the involvement of cyclists. In our predictive analysis targeting areas with elevated traffic accident risks, we employed a dichotomous methodology to delineate areas as 'high' or 'low' risk. It is critical to acknowledge the inherent

limitations of this approach. Firstly, such a binary classification inevitably results in the loss of granular data and introduces a degree of bias. This method obscures detailed nuances and variations within the dataset that could be instrumental for more in-depth analyses. However, it should be noted that this approach aligned with our primary objective, which was to pinpoint the areas with the most significant risk, rather than to explore more nuanced, intermediate risk levels. Any analysis with an aim to capture these subtler gradients in risk would necessitate a departure from this binary framework, given its propensity to oversimplify the multifaceted nature of accident risks. Furthermore, the selection of a specific threshold for risk categorization is subject to variation and can markedly affect the analysis outcomes. As previously mentioned, our chosen thresholds were informed by risk-matrix methodologies, reflecting a tailored approach to our specific analytical objectives [198, 207]. Another possible issue is related to the misinterpretation of the results. Stakeholders might misinterpret the binary categorization, especially in areas classified as low risk. Low risk does not equate to no risk, and such areas might still require attention and resources to ensure safety. The rationale for concentrating exclusively on areas with the highest risk is derived from the stipulations in Biciplans and SUMP, which explicitly mandate the identification and prioritization of areas necessitating urgent intervention [55, 58]. This approach underscores the principle of allocating resources and efforts to locations where they can have the most immediate and significant impact, in line with strategic urban planning and sustainable mobility objectives. We utilized classic machine learning models such as Support Vector Machine, Logistic Regression and Random Forest Classifier. These models were constrained by a limited selection of features and challenges such as a restricted dataset size and class imbalance. The models' performance in terms of accuracy, recall, and precision was not consistently high. Consequently, we opted for a model that optimally reduced the incidence of false negatives while maintaining an equilibrium between precision and recall. This approach acknowledges the trade-offs inherent in model selection, prioritizing a balance of key evaluation metrics. Future research aiming to enhance this, should employ more sophisticated geospatial analysis models that integrate a broader range of features and data. It would be beneficial to consider advanced methodologies, as exemplified in certain notable works such as [51, 118, 153, 216], which demonstrate the effective use of comprehensive datasets and multifaceted analytical techniques. This approach would likely yield more precise and reliable insights into geospatial patterns and risk factors. More generally, the existing scripts and models may have inherent biases or limitations. More comprehensive tools that includes a wider range of variables could be explored in future research. For example, it could be useful to create a more sophisticated, user-friendly, automated system able to work with larger areas/datasets and provide results and visualization related to bigger regions or countries in near-time as data changes. The current scripts are good for single cases but they are to a certain extent computationally expensive as they increase the area of interest and the corresponding amount of data. This study illuminates the intricacies and challenges in evaluating bicycle traffic stress levels, bikeability, and safe access to key destinations. Despite advancements, the dependence on OpenStreetMap data and the necessity to make assumptions due to data gaps, along with procedural limitations, highlight the need for more versatile and encompassing methodologies. Future research should pivot towards sophisticated models and techniques for deeper, more accurate insights, thereby enhancing urban cycling infrastructure and safety. This task presents both a technical challenge and a chance to promote safer, more accessible, and sustainable urban transport solutions.

Conclusions

The research presented in this work offers an in-depth exploration of urban cycling networks, focusing on level of traffic stress, infrastructure, safety, accessibility and their intersections with urban accidents. The *LTS Bike Plan*, developed through this research, represents a significant stride towards redefining urban cycling networks in the context of Trento and Bolzano. The *LTS Bike Plan* not only echoes the motivations and objectives laid out at the onset but also solidly answers the research questions posited, offering a comprehensive view of urban cycling dynamics. Indeed, the main goal was to dissect the complexities of these networks and understand their impact on cyclists' safety and decision-making. Significant findings arose from the meticulous identification and clustering of traffic stress areas, illuminating distinct patterns of varying stress levels across urban landscapes. Specifically, we discovered that intermediate levels of traffic stress (LTS 3) are predominant in both cities, indicating significant barriers to safe and comfortable cycling, particularly for less experienced riders. High traffic stress areas (LTS 4) are less present and typically associated with dense traffic, industrial activities, and challenging topography. This suggests a need for targeted improvements in cycling infrastructure beyond city centers. The analysis of bikeability, using the Bicycle Network Analysis (BNA) scores, revealed that central areas in both cities offer better conditions for cyclists and more access to common destinations, with challenges persisting in the peripheries. This underscores the importance of holistic urban planning that addresses the entire cityscape to enhance overall cycling conditions. This comprehensive assessment underscores the critical role of strategic urban planning in not only improving cycling infrastructure but also in fostering a network of low-stress, cyclist-friendly routes. Additionally, the focused identification of high-stress gaps within the street network emerged as a pivotal aspect of the study. By pinpointing these specific areas, the research provides a clear roadmap for government and policy interventions, directing attention and resources to the most critical zones, aligning with the demands that have emerged from the various European documents such as SUMP [163, 222, 225]. The comparison of these gaps with existing cycling plans in Trento and Bolzano showed varying degrees of overlap, indicating ongoing efforts to address these challenges. This prioritization is indeed instrumental in formulating targeted strategies that aim to significantly enhance the safety and accessibility of cycling, thereby contributing to the broader vision of sustainable and active urban mobility. Furthermore, the spatial correlation between accidents and high-stress gaps also accentuates the critical need for targeted interventions and the subsequent predictive analysis of high-risk areas goes into the same direction. This analysis, underpinned by quantitative and qualitative measures, has enabled the anticipation of potential accident hotspots, which occur mainly in areas with denser or more central street networks, thereby advocating for preemptive safety measures required by European legislations [163, 222, 225]. This result suggests that despite being less stressful, these central and dense areas do not necessarily correlate with reduced accident rates. The Plan, therefore, stands as an evidence to the potential of data-driven, policy-aligned tools in helping to transform urban landscapes into safer, more inviting, and sustainable environments for cyclists. The methodical approach,

spanning from the adaptation of the LTS scoring process to the final analysis steps, has not only validated the conceptual framework but also underscored the feasibility and effectiveness of the proposed solutions. The study's focus on specific case studies, while maintaining a broader applicability, ensures that the findings are not just contextually relevant but also scalable to other urban settings. The study faced several limitations, primarily due to its reliance on OpenStreetMap data, which can vary in accuracy and completeness, particularly in less-documented areas. This issue significantly impacts the robustness of the results, especially since the Bike Network Analysis (BNA) score heavily depends on this data. Another key limitation was the absence of certain data, notably road traffic flow data, which could have provided deeper insights into areas needing traffic stress reduction and safety improvements. Methodological constraints included making assumptions in adapting the Level of Traffic Stress (LTS) process to Italian contexts due to data gaps, particularly in road characteristics like speed limits or lane numbers. The study's approach to estimating bicycle traffic flow also involved certain presumptions, potentially limiting the accuracy in reflecting actual cycling patterns. Additionally, the use of generalized population data and the limitations in the machine learning models used for predictive analysis of high-risk areas posed challenges in achieving precision. In conclusion, despite the limits highlighted, this study is a first attempt to demonstrate the complexity of urban cycling networks and the multifaceted nature of promoting safe and accessible cycling by providing a comprehensive tool. The findings highlight the importance of continuous evaluation and adaptation of urban mobility plans to address the dynamic needs of urban cyclists. By focusing on reducing traffic stress, improving infrastructure, and enhancing safety, cities can move towards more sustainable, healthy, and inclusive urban environments. The insights from this research provide a foundation for policymakers and urban planners to develop more effective strategies for promoting cycling as a key component of urban transit systems.

Bibliography

- [1] Lorena Abad and Lucas Van der Meer. “Quantifying bicycle network connectivity in Lisbon using open data”. In: *Information* 9.11 (2018), p. 287.
- [2] Lorena Cristina Abad Crespo. “Validating a bike network analysis score based on open data as a connectivity measure of urban cycling infrastructure adapted for European Cities”. MA thesis. University of Münster, 2019.
- [3] Zainab Abbas et al. “Real-time traffic jam detection and congestion reduction using streaming graph analytics”. In: *2020 IEEE International Conference on Big Data (Big Data)*. IEEE. 2020, pp. 3109–3118.
- [4] *ADFC-Fahrradklima-Test*. <https://www.fahrradklima-test.de/>. Accessed: 11-01-2024. 2018.
- [5] Alto Adige. *In città aumenta l’uso dell’auto*. Online. Accessed: January 2024. 2023. URL: <https://www.altoadige.it/cronaca/bolzano/in-citt%C3%A0-aumenta-l-uso-dell-auto-1.1524357>.
- [6] Ravindra K Ahuja et al. “Faster algorithms for the shortest path problem”. In: *Journal of the ACM (JACM)* 37.2 (1990), pp. 213–223.
- [7] Gulsah Akar, Nicholas Fischer, and Mi Namgung. “Bicycling choice and gender case study: The Ohio State University”. In: *International journal of sustainable transportation* 7.5 (2013), pp. 347–365.
- [8] SMA Bin Al Islam and Ali Hajbabaie. “Distributed coordinated signal timing optimization in connected transportation networks”. In: *Transportation Research Part C: Emerging Technologies* 80 (2017), pp. 272–285.
- [9] American Psychological Association. *Stress*. Accessed: 15-01-2024. 2023. URL: <https://www.apa.org/topics/stress/>.
- [10] Michael Andersen. “Here’s a New Street-Level Analysis of the Biking Networks in 299 U.S. Cities”. In: *Streetsblog USA* (June 2017). Accessed: 03-07-2023. URL: <https://usa.streetsblog.org/2017/06/07/heres-a-new-street-level-analysis-of-the-biking-networks-in-299-u-s-cities>.
- [11] Mihael Ankerst et al. “OPTICS: Ordering points to identify the clustering structure”. In: *ACM Sigmod record* 28.2 (1999), pp. 49–60.
- [12] Luc Anselin. “Local indicators of spatial association—LISA”. In: *Geographical analysis* 27.2 (1995), pp. 93–115.
- [13] *Appendice 2 - Quadro generale Biciplan metropolitano*. https://www.cittametropolitana.fi.it/wp-content/uploads/Appendice-2_Quadro-generale-Biciplan-metropolitano.pdf. Accessed: 10-05-2023.

- [14] Julián Arellana et al. “Developing an urban bikeability index for different types of cyclists as a tool to prioritise bicycle infrastructure investments”. In: *Transportation Research Part A: Policy and Practice* 139 (2020), pp. 310–334.
- [15] FS Ayachi, Jonathan Dorey, and Catherine Guastavino. “Identifying factors of bicycle comfort: An online survey with enthusiast cyclists”. In: *Applied ergonomics* 46 (2015), pp. 124–136.
- [16] HM Abdul Aziz et al. “Exploring the impact of walk–bike infrastructure, safety perception, and built-environment on active transportation mode choice: a random parameter model using New York City commuter data”. In: *Transportation* 45 (2018), pp. 1207–1229.
- [17] Christopher Barrington-Leigh and Adam Millard-Ball. “The world’s user-generated road map is more than 80% complete”. In: *PloS one* 12.8 (2017), e0180698.
- [18] Marc Barthélemy. “Spatial networks”. In: *Physics reports* 499.1-3 (2011), pp. 1–101.
- [19] Gustavo EAPA Batista, Ronaldo C Prati, and Maria Carolina Monard. “A study of the behavior of several methods for balancing machine learning training data”. In: *ACM SIGKDD explorations newsletter* 6.1 (2004), pp. 20–29.
- [20] David Berrigan, Linda W Pickle, and Jennifer Dill. “Associations between street connectivity and active transportation”. In: *International journal of health geographics* 9 (2010), pp. 1–18.
- [21] *Biciplan Metropolitano*. https://pumbologna.it/Biciplan_metropolitano. Accessed: 10-05-2023.
- [22] Alexander Bigazzi et al. “Physiological markers of traffic-related stress during active travel”. In: *Transportation research part F: traffic psychology and behaviour* 84 (2022), pp. 223–238.
- [23] Norman Biggs, E Keith Lloyd, and Robin J Wilson. *Graph Theory, 1736-1936*. Oxford University Press, 1986.
- [24] Bike Ottawa. *BikeOttawa Stress Model*. GitHub repository. 2018. URL: <https://github.com/BikeOttawa/stressmodel>.
- [25] *Bike Score Methodology*. <https://www.walkscore.com/bike-score-methodology.shtml>. Accessed: December 2023.
- [26] *Bikeability*. <https://www.merriam-webster.com/dictionary/bikeability>. Accessed: 2023-12-10.
- [27] *Bikeable*. <https://en.oxforddictionaries.com/definition/bikeable>. Accessed: 2023-12-10.
- [28] Geneviève Boisjoly, Ugo Lachapelle, and Ahmed El-Geneidy. “Bicycle network performance: Assessing the directness of bicycle facilities through connectivity measures, a Montreal, Canada case study”. In: *International journal of sustainable transportation* 14.8 (2020), pp. 620–634.
- [29] *Bolzano*. <https://it.wikipedia.org/wiki/Bolzano>. Accessed: January 2024. 2024.
- [30] Phillip Bonacich. “Power and centrality: A family of measures”. In: *American journal of sociology* 92.5 (1987), pp. 1170–1182.
- [31] John Adrian Bondy and Uppaluri Siva Ramachandra Murty. *Graph theory*. Springer Publishing Company, Incorporated, 2008.
- [32] Madeline Bonsma-Fisher. *LTS-OSM*. <https://github.com/mbonsma/LTS-OSM>. 2022.

-
- [33] Soufiane Boufous et al. “Facilitators and barriers to cycling in older residents of New South Wales, Australia”. In: *Journal of Transport & Health* 21 (2021), p. 101056.
- [34] Enrico Bozzo and Massimo Franceschet. “Resistance distance, closeness, and betweenness”. In: *Social Networks* 35.3 (2013), pp. 460–469.
- [35] Ulrik Brandes. *Network analysis: methodological foundations*. Vol. 3418. Springer Science & Business Media, 2005.
- [36] Joseph Broach, Jennifer Dill, and John Gliebe. “Where do cyclists ride? A route choice model developed with revealed preference GPS data”. In: *Transportation Research Part A: Policy and Practice* 46.10 (2012), pp. 1730–1740.
- [37] Jason Brownlee. *Logistic Regression for Machine Learning*. Machine Learning Mastery. 2023. URL: <https://machinelearningmastery.com/logistic-regression-for-machine-learning>.
- [38] Ralph Buehler and Jennifer Dill. “Bikeway networks: A review of effects on cycling”. In: *Transport reviews* 36.1 (2016), pp. 9–27.
- [39] Ralph Buehler and John Pucher. “Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes”. In: *Transportation* 39 (2012), pp. 409–432.
- [40] Laura Cabral and Amy M Kim. “An empirical reappraisal of the level of traffic stress framework for segments”. In: *Travel behaviour and society* 26 (2022), pp. 143–158.
- [41] Laura Cabral, Amy M Kim, and Manish Shirgaokar. “Low-stress bicycling connectivity: Assessment of the network build-out in Edmonton, Canada”. In: *Case studies on transport policy* 7.2 (2019), pp. 230–238.
- [42] Tadeusz Caliński and Jerzy Harabasz. “A dendrite method for cluster analysis”. In: *Communications in Statistics-theory and Methods* 3.1 (1974), pp. 1–27.
- [43] Yijie Cao and Dan Shen. “Contribution of shared bikes to carbon dioxide emission reduction and the economy in Beijing”. In: *Sustainable Cities and Society* 51 (2019), p. 101749.
- [44] Denise Capasso Da Silva, David A King, and Shea Lemar. “Accessibility in practice: 20-minute city as a sustainability planning goal”. In: *Sustainability* 12.1 (2020), p. 129.
- [45] AS Cardamone, Laura Eboli, and Gabriella Mazzulla. “Drivers’ road accident risk perception. A comparison between face-to-face interview and web-based survey.” In: *Advances in Transportation Studies* 33 (2014).
- [46] Germán A Carvajal et al. “Bicycle safety in Bogotá: A seven-year analysis of bicyclists’ collisions and fatalities”. In: *Accident Analysis & Prevention* 144 (2020), p. 105596.
- [47] Ugo N Castañon and Paulo JG Ribeiro. “Bikeability and emerging phenomena in cycling: Exploratory analysis and review”. In: *Sustainability* 13.4 (2021), p. 2394.
- [48] Alvaro Caviedes and Miguel Figliozzi. “Modeling the impact of traffic conditions and bicycle facilities on cyclists’ on-road stress levels”. In: *Transportation research part F: traffic psychology and behaviour* 58 (2018), pp. 488–499.
- [49] Robert Cervero, Steve Denman, and Ying Jin. “Network design, built and natural environments, and bicycle commuting: Evidence from British cities and towns”. In: *Transport policy* 74 (2019), pp. 153–164.
- [50] Nitesh V Chawla et al. “SMOTE: synthetic minority over-sampling technique”. In: *Journal of artificial intelligence research* 16 (2002), pp. 321–357.

- [51] Chen Chen et al. “How bicycle level of traffic stress correlate with reported cyclist accidents injury severities: A geospatial and mixed logit analysis”. In: *Accident Analysis & Prevention* 108 (2017), pp. 234–244.
- [52] Guoqing Chen and Haizhong Qian. “Extracting Skeleton Lines from Building Footprints by Integration of Vector and Raster Data”. In: *ISPRS International Journal of Geo-Information* 11.9 (2022), p. 480.
- [53] Yang Chen and Yuanchang Deng. “Traffic accident risk factor identification based on complex network”. In: *IOP conference series: earth and environmental science*. Vol. 719. 3. IOP Publishing, 2021, p. 032074.
- [54] Marco Chistè. “Le piste ciclabili in città: molti gli aspetti da chiarire”. In: *L’Adige* (Aug. 2023). Accessed: December 2023. URL: <https://www.ladige.it/territori/trento/2023/08/02/le-piste-ciclabili-in-citta-molti-gli-aspetti-da-chiarire-1.3555951>.
- [55] Città di Bolzano. *Piano della Mobilità Ciclistica*. <https://opencity.comune.bolzano.it/Documenti-e-dati/Progetti-studi-e-ricerche/Piano-della-mobilita-ciclistica>. Accessed: 13-01-2024. 2022.
- [56] Sheldon Cohen, Peter J Gianaros, and Stephen B Manuck. “A stage model of stress and disease”. In: *Perspectives on Psychological Science* 11.4 (2016), pp. 456–463.
- [57] EC-European Commission et al. “Roadmap to a Single European Transport Area-Towards a competitive and resource efficient transport system”. In: *White Paper, Communication* 144 (2011).
- [58] Comune di Trento. *Biciplan*. <https://www.comune.trento.it/Aree-tematiche/Ambiente-e-territorio/Mobilita-e-traffico-urbano/Pums/Fase-III/Elaborati-grafici/Biciplan>. Accessed: 13-01-2024. 2022.
- [59] Comune di Trento. *Piano Urbano della Mobilità Sostenibile (PUMS) - Fase III: Elaborati Descrittivi*. <https://www.comune.trento.it/Aree-tematiche/Ambiente-e-territorio/Mobilita-e-traffico-urbano/Pums/Fase-III/Elaborati-descrittivi>. Accessed: 03-12-2023. 2023.
- [60] *Comuni in prov. di TN per altitudine*. <https://www.tuttitalia.it>. Accessed: December, 2023. Feb. 2012.
- [61] *Comuniciclabili FIAB: Consegnate altre 44 bandiere gialle, la rete continua a crescere*. <https://fiabitalia.it/comuniciclabili-fiab-consegnate-altre-44-bandiere-gialle-le-rete-continua-a-crescere/>. Accessed: December 2023. 2023.
- [62] Corinna Cortes and Vladimir Vapnik. “Support-vector networks”. In: *Machine learning* 20 (1995), pp. 273–297.
- [63] Alexandra D Crosswell and Kimberly G Lockwood. “Best practices for stress measurement: How to measure psychological stress in health research”. In: *Health psychology open* 7.2 (2020), p. 2055102920933072.
- [64] Record CROW. “25: Design Manual for Bicycle Traffic”. In: *CRow, The Netherlands* (2006).
- [65] Paolo Crucitti, Vito Latora, and Sergio Porta. “Centrality in networks of urban streets”. In: *Chaos: an interdisciplinary journal of nonlinear science* 16.1 (2006).
- [66] Viktor Danchuk, Olena Bakulich, and Vitaliy Svatko. “An improvement in ant algorithm method for optimizing a transport route with regard to traffic flow”. In: *Procedia Engineering* 187 (2017), pp. 425–434.

- [67] Nikola Davidovic et al. “Tagging in volunteered geographic information: an analysis of tagging practices for cities and urban regions in OpenStreetMap”. In: *ISPRS International Journal of Geo-Information* 5.12 (2016), p. 232.
- [68] David L Davies and Donald W Bouldin. “A cluster separation measure”. In: *IEEE transactions on pattern analysis and machine intelligence* 2 (1979), pp. 224–227.
- [69] William Jeffrey Davis. “Bicycle safety evaluation”. PhD thesis. Auburn University, 1987.
- [70] Delaware Valley Regional Planning Commission. *Bicycle Level of Traffic Stress (LTS) and Connectivity Analysis*. <https://www.dvrpc.org/webmaps/bike-lts/>. 2021.
- [71] Delaware Valley Regional Planning Commission. *Bike Stress Documentation: Phase 3*. Tech. rep. Accessed: 01-08-2023. Delaware Valley Regional Planning Commission, 2021.
- [72] Sybil Derrible and Christopher Kennedy. “Applications of graph theory and network science to transit network design”. In: *Transport reviews* 31.4 (2011), pp. 495–519.
- [73] Edsger W Dijkstra. “A note on two problems in connexion with graphs”. In: *Numerische mathematik* 1.1 (1959), pp. 269–271.
- [74] Jennifer Dill. “Bicycling for transportation and health: the role of infrastructure”. In: *Journal of public health policy* 30 (2009), S95–S110.
- [75] Jennifer Dill and Nathan McNeil. “Four types of cyclists? Examination of typology for better understanding of bicycling behavior and potential”. In: *Transportation Research Record* 2387.1 (2013), pp. 129–138.
- [76] Nejdet Dogru and Abdulhamit Subasi. “Traffic accident detection using random forest classifier”. In: *2018 15th learning and technology conference (L&T)*. IEEE. 2018, pp. 40–45.
- [77] Ni Dong, Helai Huang, and Liang Zheng. “Support vector machine in crash prediction at the level of traffic analysis zones: assessing the spatial proximity effects”. In: *Accident Analysis & Prevention* 82 (2015), pp. 192–198.
- [78] Richard Dowling et al. “Multimodal level of service for urban streets”. In: *Transportation Research Record* 2071.1 (2008), pp. 1–7.
- [79] Juliet Eldred. “Modeling Bicycle Comfrot with Conveyal Analysis”. In: *Medium* (2020). Accessed: 05-05-2023. URL: <https://medium.com/conveyal-blog/bike-lts-with-single-point-analysis-in-conveyal-55eecff8c0c7>.
- [80] Nikolaos Eliou, Athanasios Galanis, and Apostolos Proios. “Evaluation of the bikeability of a Greek city: Case study “City of Volos””. In: *WSEAS Transactions on Environment and Development* 5.8 (2009), pp. 545–544.
- [81] Martin Ester et al. “A density-based algorithm for discovering clusters in large spatial databases with noise”. In: *kdd*. Vol. 96. 34. 1996, pp. 226–231.
- [82] Ernesto Estrada and Philip A Knight. *A first course in network theory*. Oxford University Press, USA, 2015.
- [83] Steven Farber and Maria Grandez Marino. “Transit accessibility, land development and socioeconomic priority: A typology of planned station catchment areas in the Greater Toronto and Hamilton Area”. In: *Journal of Transport and Land Use* 10.1 (2017), pp. 879–902.
- [84] Federazione Italiana Amici della Bicicletta (FIAB). *Il Biciplan comunale*. https://www.fiab.info/download/FIAB_scheda_div_1.pdf. Accessed: 13-12-2023.

- [85] Nicholas N Ferencak and Wesley E Marshall. “Validation of bicycle level of traffic stress and perceived safety for children”. In: *Transportation research record* 2674.4 (2020), pp. 397–406.
- [86] Dillon T Fitch, James Sharpnack, and Susan L Handy. “Psychological stress of bicycling with traffic: examining heart rate variability of bicyclists in natural urban environments”. In: *Transportation research part F: traffic psychology and behaviour* 70 (2020), pp. 81–97.
- [87] Robert W Floyd. “Algorithm 97: shortest path”. In: *Communications of the ACM* 5.6 (1962), p. 345.
- [88] Linton C Freeman. “A set of measures of centrality based on betweenness”. In: *Sociometry* (1977), pp. 35–41.
- [89] Peter G Furth. “Bicycling infrastructure for all”. In: *Cycling for sustainable cities* (2021), pp. 81–100.
- [90] Peter G Furth, Maaza C Mekuria, and Hilary Nixon. “Network connectivity for low-stress bicycling”. In: *Transportation research record* 2587.1 (2016), pp. 41–49.
- [91] Peter G Furth, Bitu Sadeghinassr, and Luis Miranda-Moreno. “Slope stress criteria as a complement to traffic stress criteria, and impact on high comfort bicycle accessibility”. In: *Journal of Transport Geography* 112 (2023), p. 103708.
- [92] Gilbert C Gee and David T Takeuchi. “Traffic stress, vehicular burden and well-being: a multilevel analysis”. In: *Social science & medicine* 59.2 (2004), pp. 405–414.
- [93] M.S.D.B. Gehring. *Bikeability—Index für Dresden—Wie Fahrradfreundlich ist Dresden?* <https://heigit.org/analyzing-bikeability/>. 2019.
- [94] Jérémy Gelb and Philippe Apparicio. “Cyclists’ exposure to atmospheric and noise pollution: a systematic literature review”. In: *Transport Reviews* 41.6 (2021), pp. 742–765.
- [95] R. Geller. *Four Types of Cyclists*. Portland, 2006.
- [96] *Giretto d’Italia – bike to work 2021: Padova campione di ciclabilità urbana*. <https://www.legambiente.it/comunicati-stampa/giretto-ditalia-bike-to-work-2021-padova-campione-di-ciclabilita-urbana/>. Accessed: December 2023. 2021.
- [97] *Global Gateway*. https://international-partnerships.ec.europa.eu/policies/global-gateway_en. Accessed: 13-12-2023. 2023.
- [98] Michael F Goodchild and Linna Li. “Assuring the quality of volunteered geographic information”. In: *Spatial statistics* 1 (2012), pp. 110–120.
- [99] *GoogleMaps*. <https://https://www.google.com/maps/>.
- [100] Thomas Götschi, Jan Garrard, and Billie Giles-Corti. “Cycling as a part of daily life: a review of health perspectives”. In: *Transport Reviews* 36.1 (2016), pp. 45–71.
- [101] Elena Grigore et al. “Bikeability in Basel”. In: *Transportation Research Record: Journal of the Transportation Research Board* 2673.6 (June 2019), pp. 607–617.
- [102] R de Groot. *Design manual for bicycle traffic*. 25. 2007.
- [103] *Guida per Progettare la Ciclabilità Sicura*. https://www.fiab.info/download/2020_12_CSFIAB_GuidaProgettareciclabilitasicura.pdf. Accessed: 10-05-2023.
- [104] *H3: Uber’s Hexagonal Hierarchical Spatial Index*. <https://eng.uber.com/h3/>. Accessed: August 2023.
- [105] *H3Geo Documentation*. <https://h3geo.org/docs/>. Accessed: May 2023.

-
- [106] Mordechai Haklay and Patrick Weber. “Openstreetmap: User-generated street maps”. In: *IEEE Pervasive computing* 7.4 (2008), pp. 12–18.
- [107] Tmmit H Halefom et al. “How much traffic stress can cyclists endure?” In: *Case studies on transport policy* 10.4 (2022), pp. 2251–2261.
- [108] Hui Han, Wen-Yuan Wang, and Bing-Huan Mao. “Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning”. In: *International conference on intelligent computing*. Springer. 2005, pp. 878–887.
- [109] Handshake Cycling. *CROW-Fietsberaad Knowledge Base*. <https://handshakecycling.eu/resources/crow-fietsberaad-knowledge-base>. Accessed: 01-12-2023. 2023.
- [110] Susan Handy, Xinyu Cao, and Patricia L Mokhtarian. “Self-selection in the relationship between the built environment and walking: Empirical evidence from Northern California”. In: *Journal of the American planning association* 72.1 (2006), pp. 55–74.
- [111] Susan L Handy and Yan Xing. “Factors correlated with bicycle commuting: A study in six small US cities”. In: *International Journal of Sustainable Transportation* 5.2 (2011), pp. 91–110.
- [112] Walter G Hansen. “How accessibility shapes land use”. In: *Journal of the American Institute of planners* 25.2 (1959), pp. 73–76.
- [113] Ghanim Al-Hasani, Md Asaduzzaman, and Abdel-Hamid Soliman. “Comparison of spatial regression models with road traffic accidents data”. In: *Proceedings of the International Conference on Statistics: Theory and Applications (ICSTA '19)*. ICSTA. 2019.
- [114] Haibo He et al. “ADASYN: Adaptive synthetic sampling approach for imbalanced learning”. In: *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*. Ieee. 2008, pp. 1322–1328.
- [115] Benjamin Herfort et al. “A spatio-temporal analysis investigating completeness and inequalities of global urban building data in OpenStreetMap”. In: *Nature Communications* 14.1 (2023), p. 3985.
- [116] Robin Hickman and David Banister. *Transport, climate change and the city*. Routledge, 2014.
- [117] Simon Hoedl, Sylvia Titze, and Pekka Oja. “The bikeability and walkability evaluation table: Reliability and application”. In: *American journal of preventive medicine* 39.5 (2010), pp. 457–459.
- [118] Jorge A Huertas et al. “Level of traffic stress-based classification: a clustering approach for Bogotá, Colombia”. In: *Transportation Research Part D: Transport and Environment* 85 (2020), pp. 102–420.
- [119] Angela Hull and Craig O’holleran. “Bicycle infrastructure: can good design encourage cycling?” In: *Urban, Planning and Transport Research* 2.1 (2014), pp. 369–406.
- [120] *Humanitarian Data Exchange*. <https://data.humdata.org/>. Accessed: December 2023.
- [121] Tetsuro Hyodo, Norikazu Suzuki, and Katsumi Takahashi. “Modeling of bicycle route and destination choice behavior for bicycle road network plan”. In: *Transportation Research Record* 1705.1 (2000), pp. 70–76.
- [122] Michael Iacono, Kevin J Krizek, and Ahmed El-Geneidy. “Measuring non-motorized accessibility: issues, alternatives, and execution”. In: *Journal of Transport Geography* 18.1 (2010), pp. 133–140.

- [123] A Ilie et al. “The use of the bicycle compatibility index in identifying gaps and deficiencies in bicycle networks”. In: *IOP Conference Series: Materials Science and Engineering*. Vol. 161. 1. IOP Publishing. 2016, p. 012097.
- [124] Ahmadreza Faghieh Imani, Eric J Miller, and Shoshanna Saxe. “Cycle accessibility and level of traffic stress: A case study of Toronto”. In: *Journal of transport geography* 80 (2019), p. 102496.
- [125] *Incidenti Open Data - Trentino*. <https://dati.trentino.it/dataset/incidenti-open-data>. Accessed: November 2023.
- [126] ISTAT. *Bilancio demografico mensile anno 2023 (dati provvisori)*. <https://demo.istat.it>. Accessed: 18-12-2023. Dec. 2023.
- [127] ISTAT. *Bilancio demografico mensile anno 2023 (dati provvisori)*. Online. Accessed: 28-12-2023. Dec. 2023. URL: <https://demo.istat.it>.
- [128] Søren Underlien Jensen. “Pedestrian and bicyclist level of service on roadway segments”. In: *Transportation research record* 2031.1 (2007), pp. 43–51.
- [129] Jamal Jokar Arsanjani et al. “An introduction to OpenStreetMap in Geographic Information Science: Experiences, research, and applications”. In: *OpenStreetMap in GIScience: Experiences, research, and applications* (2015), pp. 1–15.
- [130] Mohamed Bayoumi Kamel and Tarek Sayed. “The impact of bike network indicators on bike kilometers traveled and bike safety: A network theory approach”. In: *Environment and Planning B: Urban Analytics and City Science* 48.7 (2021), pp. 2055–2072.
- [131] Mohamed Bayoumi Kamel, Tarek Sayed, and Alexander Bigazzi. “A composite zonal index for biking attractiveness and safety”. In: *Accident Analysis & Prevention* 137 (Mar. 2020), p. 105439.
- [132] Lei Kang, Yingge Xiong, and Fred L Mannering. “Statistical analysis of pedestrian perceptions of sidewalk level of service in the presence of bicycles”. In: *Transportation Research Part A: Policy and Practice* 53 (2013), pp. 10–21.
- [133] Debra K Kellstedt et al. “A scoping review of bikeability assessment methods”. In: *Journal of community health* 46 (2021), pp. 211–224.
- [134] Margaret Kent and Alex Karner. “Prioritizing low-stress and equitable bicycle networks using neighborhood-based accessibility measures”. In: *International journal of sustainable transportation* 13.2 (2019), pp. 100–110.
- [135] Thomas Kluyver et al. “Jupyter Notebooks – a publishing format for reproducible computational workflows”. In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. Ed. by F. Loizides and B. Schmidt. IOS Press. 2016, pp. 87–90.
- [136] *Kontur Population Italy Dataset*. <https://data.humdata.org/dataset/kontur-population-italy>. Accessed: December 2023.
- [137] Thorsten Koska and Frederic Rudolph. “The role of walking and cycling in reducing congestion: a portfolio of measures”. In: (2017).
- [138] Patricia Jasmin Krenn, Pekka Oja, Sylvia Titze, et al. “Development of a bikeability index to assess the bicycle-friendliness of urban environments”. In: *Open Journal of Civil Engineering* 5.04 (2015), p. 451.
- [139] Patricia Jasmin Krenn, Pekka Oja, and Sylvia Titze. “Route choices of transport bicyclists: a comparison of actually used and shortest routes”. In: *International journal of behavioral nutrition and physical activity* 11.1 (2014), pp. 1–7.

-
- [140] *La Statistica ACI-Istat-Sintesi Dello Studio*. 2022. Rome, Italy: ACI, 2022.
- [141] *la Top Ten delle città italiane più bike friendly: l'Emilia Romagna primeggia*. <https://www.cyclingnotes.it/il-3-giugno-e-la-giornata-mondiale-della-bicicletta-la-top-10-delle-citta-italiane-piu-bike-friendly/>. Accessed: 15-12-2023. 2019.
- [142] Bruce W Landis. “Bicycle system performance measures: the interaction hazard and latent demand score models”. In: *ITE journal* 66.2 (1996).
- [143] Jacob Larsen and Ahmed El-Geneidy. “A travel behavior analysis of urban cycling facilities in Montréal, Canada”. In: *Transportation research part D: transport and environment* 16.2 (2011), pp. 172–177.
- [144] Jacob Larsen, Zachary Patterson, and Ahmed El-Geneidy. “Build it. But where? The use of geographic information systems in identifying locations for new cycling infrastructure”. In: *International Journal of Sustainable Transportation* 7.4 (2013), pp. 299–317.
- [145] Richard S Lazarus and Susan Folkman. *Stress, appraisal, and coping*. Springer publishing company, 1984.
- [146] Heng Li et al. “Data-driven surrogate modeling: Introducing spatial lag to consider spatial autocorrelation of flooding within urban drainage systems”. In: *Environmental Modelling & Software* 161 (2023), p. 105623.
- [147] Xiaojian Liu, Ourania Kounadi, and Raul Zurita-Milla. “Incorporating spatial autocorrelation in machine learning models using spatial lag and eigenvector spatial filtering features”. In: *ISPRS International Journal of Geo-Information* 11.4 (2022), p. 242.
- [148] Fernando López. *SHAP (SHapley Additive exPlanations)*. Towards Data Science. 2021. URL: <https://towardsdatascience.com/shap-shapley-additive-explanations-5a2a271ed9c3>.
- [149] Michael Lowry and Tracy Hadden Loh. “Quantifying bicycle network connectivity”. In: *Preventive medicine* 95 (2017), S134–S140.
- [150] Michael B Lowry, Peter Furth, and Tracy Hadden-Loh. “Prioritizing new bicycle facilities to improve low-stress network connectivity”. In: *Transportation Research Part A: Policy and Practice* 86 (2016), pp. 124–140.
- [151] Michael B Lowry et al. “Assessment of communitywide bikeability with bicycle level of service”. In: *Transportation research record* 2314.1 (2012), pp. 41–48.
- [152] Tao Lu et al. “The traffic accident hotspot prediction: Based on the logistic regression method”. In: *2015 International Conference on Transportation Information and Safety (ICTIS)*. IEEE. 2015, pp. 107–110.
- [153] Weike Lu et al. “Integrating machine learning into path analysis for quantifying behavioral pathways in bicycle-motor vehicle crashes”. In: *Accident Analysis & Prevention* 168 (2022), p. 106622.
- [154] Wesley E Marshall and Norman W Garrick. “Effect of street network design on walking and biking”. In: *Transportation Research Record* 2198.1 (2010), pp. 103–115.
- [155] Giulio Mattioli et al. “The political economy of car dependence: A systems of provision approach”. In: *Energy Research & Social Science* 66 (2020), p. 101486.
- [156] Stephen McDaniel, Michael B Lowry, and Michael Dixon. “Using origin–destination centrality to estimate directional bicycle volumes”. In: *Transportation Research Record* 2430.1 (2014), pp. 12–19.

- [157] Leland McInnes, John Healy, and Steve Astels. “hdbscan: Hierarchical density based clustering.” In: *J. Open Source Softw.* 2.11 (2017), p. 205.
- [158] Nathan McNeil. “Bikeability and the 20-min neighborhood: How infrastructure and destinations influence bicycle accessibility”. In: *Transportation research record* 2247.1 (2011), pp. 53–63.
- [159] Maaza C Mekuria, Bruce Appleyard, and Hilary Nixon. “Improving livability using green and active modes: A traffic stress level analysis of transit, bicycle, and pedestrian access and mobility”. In: (2017).
- [160] Maaza C Mekuria, Peter G Furth, and Hilary Nixon. “Low-stress bicycling and network connectivity”. In: (2012).
- [161] Microsoft Corporation. *Visual Studio Code*. <https://code.visualstudio.com/>.
- [162] Harvey J Miller. “Tobler’s first law and spatial analysis”. In: *Annals of the association of American geographers* 94.2 (2004), pp. 284–289.
- [163] Ministero delle Infrastrutture e dei Trasporti. *Linee guida per la redazione e l’attuazione del Biciplan*. Legge 2/2018, articolo 6. 2018.
- [164] Ministero delle Infrastrutture e della Mobilità Sostenibili. *Piano Generale Mobilità Ciclista*. <https://www.mit.gov.it/>. Accessed: 17-01-2024. Aug. 2022.
- [165] *Mobilità Sostenibile*. <https://www.mase.gov.it/pagina/mobilita-sostenibile>.
- [166] Montgomery County Planning Department. *Modified Level of Traffic Stress Methodology*. <https://mcatlas.org/bikestress/documentation/ModifiedLevelOfTrafficStressMethodology.pdf>. Accessed: 20-01-2024. 2021.
- [167] Peter Mooney, Marco Minghini, et al. “A review of OpenStreetMap data”. In: *Mapping and the citizen sensor* (2017), pp. 37–59.
- [168] Sarah K Moran et al. “Lowering bicycle stress one link at a time: where should we invest in infrastructure?” In: *Transportation research record* 2672.36 (2018), pp. 33–41.
- [169] Eleni Mougiakou and Yorgos N Photis. “Urban green space network evaluation and planning: Optimizing accessibility based on connectivity and raster gis analysis”. In: *European journal of geography* 5.4 (2014), pp. 19–46.
- [170] KSV Muralidhar. *What is Stratified Cross-Validation in Machine Learning*. Towards Data Science. 2021. URL: <https://towardsdatascience.com/what-is-stratified-cross-validation-in-machine-learning-8844f3e7ae8e>.
- [171] Brendan Murphy and Andrew Owen. “Implementing low-stress bicycle routing in national accessibility evaluation”. In: *Transportation Research Record* 2673.5 (2019), pp. 240–249.
- [172] Rakshita Nagalla, Prasanna Pothuganti, and Digvijay S Pawar. “Analyzing gap acceptance behavior at unsignalized intersections using support vector machines, decision tree and random forests”. In: *Procedia Computer Science* 109 (2017), pp. 474–481.
- [173] Luis Guillermo Natera Orozco et al. “Data-driven strategies for optimal bicycle network growth”. In: *Royal society open science* 7.12 (2020), p. 201130.
- [174] Mark EJ Newman. “A measure of betweenness centrality based on random walks”. In: *Social networks* 27.1 (2005), pp. 39–54.
- [175] Thomas Alexander Sick Nielsen and Hans Skov-Petersen. “Bikeability–Urban structures supporting cycling. Effects of local, urban and regional scale urban form factors on cycling from home and workplace locations in Denmark”. In: *Journal of Transport Geography* 69 (2018), pp. 36–44.

-
- [176] Pirouz Nourian et al. “Easiest paths for walking and cycling”. In: *10TH INTERNATIONAL SPACE SYNTAX SYMPOSIUM*. 2015.
- [177] Tessio Novack, Zhiyong Wang, and Alexander Zipf. “A system for generating customized pleasant pedestrian routes based on OpenStreetMap data”. In: *Sensors* 18.11 (2018), p. 3794.
- [178] Esko Nuutila and Eljas Soisalon-Soininen. “On finding the strongly connected components in a directed graph”. In: *Information processing letters* 49.1 (1994), pp. 9–14.
- [179] Pekka Oja et al. “Health benefits of cycling: a systematic review”. In: *Scandinavian journal of medicine & science in sports* 21.4 (2011), pp. 496–509.
- [180] OpenStreetMap. <https://www.openstreetmap.org/>.
- [181] OpenStreetMap contributors. *OpenStreetMap Copyright*. <https://www.openstreetmap.org/copyright>. Retrieved from OpenStreetMap. 2023.
- [182] OpenStreetMap Wiki contributors. *Stats - OpenStreetMap Wiki*. <https://wiki.openstreetmap.org/wiki/>. Retrieved from OpenStreetMap Wiki. 2023.
- [183] *OpenStreetMap: The Wikipedia of Maps [Infographic]*. <https://m2050.media/en/openstreetmap-the-wikipedia-of-maps-infographic/>. Accessed: 17-01-2024. 2023. URL: <https://m2050.media/en/openstreetmap-the-wikipedia-of-maps-infographic/>.
- [184] Ahmed Osama and Tarek Sayed. “Evaluating the impact of bike network indicators on cyclist safety using macro-level collision prediction models”. In: *Accident Analysis & Prevention* 97 (2016), pp. 28–37.
- [185] Andrew Owen and David M Levinson. “Modeling the commute mode share of transit using continuous accessibility to jobs”. In: *Transportation research part A: policy and practice* 74 (2015), pp. 110–122.
- [186] Aakash Parmar, Rakesh Katariya, and Vatsal Patel. “A review on random forest: An ensemble classifier”. In: *International conference on intelligent data communication technologies and internet of things (ICICI) 2018*. Springer. 2019, pp. 758–763.
- [187] Aritz Parra and Associated Press. “Cities Are Becoming More Cycle-Friendly as They Emerge From Lockdown”. In: *AFAR* (May 2020). URL: <https://www.afar.com/magazine/cities-are-becoming-more-cycle-friendly-as-they-emerge-from-lockdown>.
- [188] Karl Pearson. “X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 50.302 (July 1900), pp. 157–175. DOI: [10.1080/14786440009463897](https://doi.org/10.1080/14786440009463897). URL: <https://doi.org/10.1080/14786440009463897>.
- [189] PeopleForBikes. *Bicycle Network Analysis*. <https://bna.peopleforbikes.org/>. Retrieved from PeopleForBikes website. 2019.
- [190] PeopleForBikes. *City Ratings*. <https://cityratings.peopleforbikes.org/ratings>. Accessed: 01-01-2024. 2024.
- [191] PeopleForBikes. *Methodology BNA*. <https://cityratings.peopleforbikes.org/about/methodology>. Retrieved from PeopleForBikes website. 2017.
- [192] Jonathan D Phillips, Wolfgang Schwanghart, and Tobias Heckmann. “Graph theory in the geosciences”. In: *Earth-Science Reviews* 143 (2015), pp. 147–160.

BIBLIOGRAPHY

- [193] Laura Po et al. “From sensors data to urban traffic flow analysis”. In: *2019 IEEE International Smart Cities Conference (ISC2)*. IEEE. 2019, pp. 478–485.
- [194] Sergio Porta, Paolo Crucitti, and Vito Latora. “The network analysis of urban streets: a primal approach”. In: *Environment and Planning B: planning and design* 33.5 (2006), pp. 705–725.
- [195] Anna K Porter et al. “Bikeability: assessing the objectively measured environment in relation to recreation and transportation bicycling”. In: *Environment and Behavior* 52.8 (2020), pp. 861–894.
- [196] Gabriele Prati, Luca Pietrantoni, and Federico Fraboni. “Using data mining techniques to predict the severity of bicycle crashes”. In: *Accident Analysis & Prevention* 101 (2017), pp. 44–54.
- [197] Gabriele Prati et al. “Gender differences in cyclists’ crashes: an analysis of routinely recorded crash data”. In: *International journal of injury control and safety promotion* 26.4 (2019), pp. 391–398.
- [198] Project Management Institute. *A Guide to the Project Management Body of Knowledge (PMBOK Guide) – Sixth Edition*. Table 11-1, Page 407. Pennsylvania, USA: Project Management Institute, 2017.
- [199] John Pucher and Ralph Buehler. “Cycling towards a more sustainable transport future”. In: *Transport reviews* 37.6 (2017), pp. 689–694.
- [200] John Pucher and Lewis Dijkstra. “Making walking and cycling safer: lessons from Europe”. In: *Transportation Quarterly* 54.3 (2000), pp. 25–50.
- [201] Rami Puzis et al. “Augmented betweenness centrality for environmentally aware traffic monitoring in transportation networks”. In: *Journal of Intelligent Transportation Systems* 17.1 (2013), pp. 91–105.
- [202] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Padova, Italy, 2023. URL: <https://www.r-project.org/>.
- [203] Kelcie Ralph and Leigh Ann Von Hagen. “Will parents let their children bike on “low stress” streets? Validating level of traffic stress for biking”. In: *Transportation research part F: traffic psychology and behaviour* 65 (2019), pp. 280–291.
- [204] Giulia Reggiani et al. “Understanding bikeability: a methodology to assess urban networks”. In: *Transportation* 49.3 (2022), pp. 897–925.
- [205] Anna-Karina Reibold. “How the Humble Bicycle is Transforming Cities”. In: *European Cyclists’ Federation* (Nov. 2020). Accessed: 01-11-2024. URL: <https://ecf.com/news-and-events/news/how-humble-bicycle-transforming-cities>.
- [206] CGS Richards et al. “Neighborhood bikeability and BMI in NYC”. In: *Active Living Research Annual Conference 2010, San Diego, CA*. 2010.
- [207] *Risk Impact/Probability Charts*. Accessed: 2024-01-27. MindTools. URL: <https://www.mindtools.com/ah8ju2z/risk-impactprobability-charts>.
- [208] Peter J Rousseeuw. “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis”. In: *Journal of computational and applied mathematics* 20 (1987), pp. 53–65.
- [209] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc. Boston, MA, 2023. URL: <http://www.rstudio.com/>.

-
- [210] Greg Rybarczyk and Changshan Wu. “Bicycle facility planning using GIS and multi-criteria decision analysis”. In: *Applied Geography* 30.2 (2010), pp. 282–293.
- [211] Tayebeh Saghapour, Sara Moridpour, and Russell G Thompson. “Measuring cycling accessibility in metropolitan areas”. In: *International journal of sustainable transportation* 11.5 (2017), pp. 381–394.
- [212] Jonas Schmid-Querg, Andreas Keler, and Georgios Grigoropoulos. “The Munich Bikeability Index: A Practical Approach for Measuring Urban Bikeability”. In: *Sustainability* 13.1 (Jan. 2021), p. 428.
- [213] Jessica E Schoner and David M Levinson. “The missing link: Bicycle infrastructure networks and ridership in 74 US cities”. In: *Transportation* 41 (2014), pp. 1187–1204.
- [214] Alexander Schrijver. “On the history of the shortest path problem”. In: *Documenta Mathematica* 17.1 (2012), pp. 155–167.
- [215] Sukhjit Singh Sehra, Jaiteg Singh, and Hardeep Singh Rai. “Assessment of OpenStreetMap data-a review”. In: *arXiv preprint arXiv:1309.6608* (2013).
- [216] SS Radiah Shariff et al. “Determining hotspots of road accidents using spatial analysis”. In: *Indonesian Journal of Electrical Engineering and Computer Science* 9.1 (2018), pp. 146–151.
- [217] Ayyoob Sharifi. “Resilient urban forms: A review of literature on streets and street networks”. In: *Building and Environment* 147 (2019), pp. 171–187.
- [218] Sixt Magazine. *Le 5 città più ciclabili d’Italia*. Online. Accessed: November 2023. 2023. URL: <https://www.sixt.it/magazine/viaggi/le-5-citta-piu-ciclabili-d-italia/>.
- [219] *sklearn.model_selection.StratifiedKfold*. scikit-learn. URL: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKfold.html.
- [220] Alex Sorton and Thomas Walsh. “Bicycle stress level as a tool to evaluate urban and suburban bicycle compatibility”. In: *Transportation Research Record* (1994), pp. 17–17.
- [221] Christoph Steinacker et al. “Demand-driven design of bicycle infrastructure networks for improved urban bikeability”. In: *Nature Computational Science* 2.10 (2022), pp. 655–664.
- [222] *SUMP Concept*. <https://www.eltis.org/mobility-plans/sump-concept>. Accessed: 13-12-2023. 2021.
- [223] Chenshuo Sun et al. “Role of road network features in the evaluation of incident impacts on urban traffic mobility”. In: *Transportation research part B: methodological* 117 (2018), pp. 101–116.
- [224] *Sustainable Development Goals & Cycling*. <https://unric.org/en/sustainable-development-goals-cycling/>. Accessed: 13-12-2023. 2023.
- [225] *Sustainable Urban Mobility Planning and Monitoring*. https://transport.ec.europa.eu/transport-themes/urban-transport/sustainable-urban-mobility-planning-and-monitoring_en. Accessed: 13-12-2023. 2013.
- [226] *Sustainable Urban Mobility Plans*. https://urban-mobility-observatory.transport.ec.europa.eu/sustainable-urban-mobility-plans_en. Accessed: 10-01-2024.
- [227] Michael Szell et al. “Growing urban bicycle networks”. In: *Scientific reports* 12.1 (2022), p. 6765.
- [228] S Tarquini et al. “TINITALY, a digital elevation model of Italy with a 10 meters cell size”. In: *Istituto Nazionale di Geofisica e Vulcanologia (INGV)* 10 (2007).

- [229] *The Copenhagenize Bicycle Friendly Cities Index*. <http://copenhagenizeindex.eu>. 2017.
- [230] *Trento*. <https://it.wikipedia.org/wiki/Trento>. Accessed: January 2024. 2024.
- [231] Hannah Twaddell et al. *Guidebook for Measuring Multimodal Network Connectivity*. Tech. rep. United States. Federal Highway Administration. Office of Planning . . . , 2018.
- [232] David S Vale, Miguel Saraiva, and Mauro Pereira. “Active accessibility: A review of operational measures of walking and cycling accessibility”. In: *Journal of transport and land use* 9.1 (2016), pp. 209–235.
- [233] Francesca Valent et al. “Risk factors for fatal road traffic accidents in Udine, Italy”. In: *Accident Analysis & Prevention* 34.1 (2002), pp. 71–84.
- [234] Guido Van Rossum, Fred L Drake, et al. *Python reference manual*. Vol. 111. Centrum voor Wiskunde en Informatica Amsterdam, 1995.
- [235] Ane Rahbek Vierø, Anastassia Vybornova, and Michael Szell. “BikeDNA: A tool for bicycle infrastructure data and network assessment”. In: *Environment and Planning B: Urban Analytics and City Science* (2023), p. 23998083231184471.
- [236] Christina Vietinghoff. “An intersectional analysis of barriers to cycling for marginalized communities in a cycling-friendly French City”. In: *Journal of transport geography* 91 (2021), p. 102967.
- [237] Anastassia Vybornova et al. “Automated detection of missing links in bicycle networks”. In: *Geographical Analysis* 55.2 (2023), pp. 239–267.
- [238] Lina Wahlgren and Peter Schantz. “Bikeability and methodological issues using the active commuting route environment scale (ACRES) in a metropolitan setting”. In: *BMC medical research methodology* 11 (2011), pp. 1–20.
- [239] Haizhong Wang et al. “Does bicycle network level of traffic stress (LTS) explain bicycle travel behavior? Mixed results from an Oregon case study”. In: *Journal of transport geography* 57 (2016), pp. 8–18.
- [240] Wencheng Wang et al. “Factors influencing traffic accident frequencies on urban roads: a spatial panel time-fixed effects error model”. In: *PLoS one* 14.4 (2019), e0214539.
- [241] *Web Feature Service of Bolzano Municipality*. <http://sit.comune.bolzano.it/geoserver/wfs?service=WFS>. Accessed: November 2023.
- [242] Meghan Winters et al. “Mapping bikeability: a spatial tool to support sustainable travel”. In: *Environment and Planning B: Planning and Design* 40.5 (2013), pp. 865–883.
- [243] Meghan Winters et al. “Motivators and deterrents of bicycling: comparing influences on decisions to ride”. In: *Transportation* 38 (2011), pp. 153–168.
- [244] Laura Wysling and Ross S Purves. “Where to improve cycling infrastructure? Assessing bicycle suitability and bikeability with open data in the city of Paris”. In: *Transportation research interdisciplinary perspectives* 15 (2022), p. 100648.
- [245] Zizhen Xu and Shauhrat S Chopra. “Interconnectedness enhances network resilience of multimodal public transportation systems for Safe-to-Fail urban mobility”. In: *Nature communications* 14.1 (2023), p. 4291.
- [246] Hong Xue et al. “Road network intersection density and childhood obesity risk in the US: a national longitudinal study”. In: *Public health* 178 (2020), pp. 31–37.

-
- [247] Kaoru Yamaoka, Yusuke Kumakoshi, and Yuji Yoshimura. “Local betweenness centrality analysis of 30 european cities”. In: *Urban Informatics and Future Cities*. Springer, 2021, pp. 527–547.
- [248] Pengyao Ye, Bo Wu, and Wenbo Fan. “Modified betweenness-based measure for prediction of traffic flow on urban roads”. In: *Transportation Research Record* 2563.1 (2016), pp. 144–150.
- [249] Mohamed Anwer Zayed. “Towards an index of city readiness for cycling”. In: *International journal of transportation science and technology* 5.3 (2016), pp. 210–225.
- [250] Dapeng Zhang, David José Ahouagi Vaz Magalhães, and Xiaokun Cara Wang. “Prioritizing bicycle paths in Belo Horizonte City, Brazil: Analysis based on user preferences and willingness considering individual heterogeneity”. In: *Transportation Research Part A: Policy and Practice* 67 (2014), pp. 268–278.
- [251] Liming Zhang and Dieter Pfoser. “Using OpenStreetMap point-of-interest data to model urban change—A feasibility study”. In: *PloS one* 14.2 (2019), e0212606.
- [252] Yuanyuan Zhang et al. “Investigating the associations between road network structure and non-motorist accidents”. In: *Journal of transport geography* 42 (2015), pp. 34–47.

BIBLIOGRAPHY

List of Figures

2.1	Graph representation. Each node is associated with a letter.	15
2.2	Example of a graph in order to show how Dijkstra’s algorithm works	17
2.3	Levels of Traffic Stress. Adapted from Alta Planning	21
3.1	Interactive Grid of Tititaly’s Downloadable Area: A Detailed View of the Geospatial Selection Interface	31
3.2	Population Density of Italy Visualized in 400m H3 Hexagons by Kontur, Courtesy of Humanitarian Data Exchange	32
3.3	Close-Up View of Traffic Accidents in the City of Trento Over the Last 5 Years	33
3.4	Diagram Illustrating the General Framework of the Analytical Tool	35
3.5	Diagram Illustrating the Level of Traffic Stress Adaptation Process	41
3.6	Diagram Illustrating the Infrastructure and Network Analysis Process	45
3.7	Diagram Illustrating the LTS - IPDC Procedure	50
3.8	Diagram Illustrating the Qualitative Analysis Process	55
3.9	Diagram Illustrating the BNA Score Process	57
3.10	Diagram Illustrating the Final Analysis Process	64
4.1	Top: Stress Network - Level of Traffic Stress in Trento. Bottom: Choropleth Map of Level of Traffic Stress in Trento.	66
4.2	Left: Distribution of Traffic Stress in Trento’s nodes. Right: Distribution of Traffic Stress in Trento’s streets.	66
4.3	Top: Stress Network - Level of Traffic Stress in Bolzano. Bottom: Choropleth Map of Level of Traffic Stress in Bolzano.	67
4.4	Left: Distribution of Traffic Stress in Bolzano’s nodes. Right: Distribution of Traffic Stress in Bolzano’s streets.	68
4.5	Trento - Street Network	69
4.6	Bolzano - Street Network	70
4.7	Top: Street Network Orientation in Trento. Bottom: Street Network Orientation - Polar Plot in Trento.	72
4.8	Top: Street Network Orientation in Bolzano. Bottom: Street Network Orientation - Polar Plot in Bolzano.	73
4.9	Top Left: Node Degree Centrality; Top Right: Node Closeness Centrality; Bottom Left: Node Betweenness Centrality; Bottom Right: Edge Betweenness Centrality — all within the Trento Network.	75
4.10	Top Left: LTS - Node Degree Centrality; Top Right: LTS - Node Closeness Centrality; Bottom Left: LTS - Node Betweenness Centrality; Bottom Right: LTS - Edge Betweenness Centrality — all within the Trento Network.	76

LIST OF FIGURES

4.11	Top: Distribution of nodes for Degree Centrality values for Trento. Bottom: Top 10 nodes for Degree Centrality values in Trento.	76
4.12	Top: Distribution of nodes for Closeness Centrality values for Trento. Bottom: Top 10 nodes for Closeness Centrality values in Trento.	77
4.13	Top: Distribution of nodes for Betweenness Centrality values for Trento. Bottom: Top 10 nodes for Betweenness Centrality values in Trento.	77
4.14	High Stress Subgraph in Trento	78
4.15	Top 5 High-Stress Connected Components in Trento	79
4.16	Hexbin map of Trento - Intersection Density	79
4.17	Kernel Density Estimation Map of Trento - Intersection Density	79
4.18	Top Left: Node Degree Centrality; Top Right: Node Closeness Centrality; Bottom Left: Node Betweenness Centrality; Bottom Right: Edge Betweenness Centrality — all within the Bolzano Network.	83
4.19	Top Left: LTS - Node Degree Centrality; Top Right: LTS - Node Closeness Centrality; Bottom Left: LTS - Node Betweenness Centrality; Bottom Right: LTS - Edge Betweenness Centrality — all within the Bolzano Network.	83
4.20	Top: Distribution of nodes for Degree Centrality values for Bolzano. Bottom: Top 10 nodes for Degree Centrality values in Bolzano.	84
4.21	Top: Distribution of nodes for Closeness Centrality values for Bolzano. Bottom: Top 10 nodes for Closeness Centrality values in Bolzano.	84
4.22	Top: Distribution of nodes for Betweenness Centrality values for Bolzano. Bottom: Top 10 nodes for Betweenness Centrality values in Bolzano.	85
4.23	High Stress Subgraph in Bolzano	86
4.24	Top 5 High-Stress Connected Components in Bolzano	86
4.25	Hexbin map of Bolzano - Intersection Density	87
4.26	Kernel Density Estimation Map of Bolzano - Intersection Density	87
4.27	Top: DBSCAN Clustering Method applied for Trento. Bottom: DBSCAN Clustering Method applied for Trento over city network layer.	89
4.28	Top: HDBSCAN Clustering Method applied for Trento. Bottom: HDBSCAN Clustering Method applied for Trento over city network layer.	90
4.29	Top: OPTICS Clustering Method applied for Trento. Bottom: OPTICS Clustering Method applied for Trento over city network layer.	92
4.30	Top: DBSCAN Clustering Method applied for Bolzano. Bottom: DBSCAN Clustering Method applied for Bolzano over city network layer.	93
4.31	Top: HDBSCAN Clustering Method applied for Bolzano. Bottom: HDBSCAN Clustering Method applied for Bolzano over city network layer.	94
4.32	Top: OPTICS Clustering Method applied for Bolzano. Bottom: OPTICS Clustering Method applied for Bolzano over city network layer.	95
4.33	Top 10 Low Stress Connected Components in Trento	96
4.34	High-Low Stress components in Trento	97
4.35	Gaps plot in Trento	97
4.36	Low Stress Contact Nodes in Trento	98
4.37	Benefit Metric and Gap Rank graph in Trento	98
4.38	Top: Declustered Gaps in Trento. Bottom: Classified Gaps in Trento	99
4.39	Top 10 Low Stress Connected Components in Bolzano	100
4.40	High-Low Stress components in Bolzano	100
4.41	Gaps plot in Bolzano	101
4.42	Low Stress Contact Nodes in Bolzano	101
4.43	Benefit Metric and Gap Rank graph in Bolzano	102

4.44	Top: Declustered Gaps in Bolzano. Bottom: Classified Gaps in Bolzano	103
4.45	Biking time to the nearest POI within 15 minutes in Trento	105
4.46	Biking time to the nearest POI within 15 minutes in Bolzano	105
4.47	Biking time to the nearest POIs (5th and 10th) within 15 minutes in Trento . . .	106
4.48	Biking time to the nearest POIs (5th and 10th) within 15 minutes in Bolzano . .	106
4.49	Population distribution in Trento H3 grid, resolution 200m. Red areas are the one more populated (more than 2% of population), followed by yellow one (more than 1% of population) and green one.	107
4.50	H3 hexagon grid of Trento colored according to BNA scores. Green colored hexagons are the ones with higher BNA scores as it can be seen from the legend on the top-right.	108
4.51	Distribution of Hexagons for Rounded BNA Scores	108
4.52	Population distribution in Bolzano H3 grid, resolution 200m. Red areas are the one more populated (more than 2% of population), followed by yellow one (more than 1% of population) and green one.	109
4.53	H3 hexagon grid of Bolzano colored according to BNA scores. Green colored hexagons are the ones with higher BNA scores as it can be seen from the legend on the top-right.	109
4.54	Distribution of Hexagons for Rounded BNA Scores in Bolzano	110
4.55	Left: Overview of Traffic Accidents in the Trento Municipality (2018-2023). Right: Heatmap of Accident Hotspots in Trento	112
4.56	Left: Number of Accidents in streets for Level of Traffic Stress in Trento. Right: Number of Accidents in intersections for Level of Traffic Stress in Trento	112
4.57	Left: Number of Accidents in streets for Type of Stress in Trento. Right: Number of Accidents in intersections for Type of Stress in Trento	112
4.58	Left: Analysis of Traffic Accidents by Speed Limit in Trento Streets. Right: Trento Street Accident Frequency related with the Number of Lanes.	113
4.59	Frequency of Accidents by Road Type in Trento	113
4.60	Distribution of accidents for different lanes and max speed values in Trento network	114
4.61	Left: Overview of Traffic Accidents in the Bolzano Municipality (2018-2023). Right: Heatmap of Accident Hotspots in Bolzano	114
4.62	Left: Number of Accidents in streets for Level of Traffic Stress in Bolzano. Right: Number of Accidents in intersections for Level of Traffic Stress in Bolzano	115
4.63	Left: Number of Accidents in streets for Type of Stress in Bolzano. Right: Number of Accidents in intersections for Type of Stress in Bolzano	115
4.64	Left: Analysis of Traffic Accidents by Speed Limit in Bolzano Streets. Right: Bolzano Street Accident Frequency related with the Number of Lanes.	115
4.65	Frequency of Accidents by Road Type in Bolzano	116
4.66	Distribution of accidents for different lanes and max speed values in Bolzano network	116
4.67	Complete High-Stress Gaps with Varied Degrees of Accident Risk Association. .	117
4.68	Declustered High-Stress Gaps with Varied Degrees of Accident Risk Association. .	118
4.69	Complete High-Stress Gaps with Varied Degrees of Accident Risk Association. .	119
4.70	Declustered High-Stress Gaps with Varied Degrees of Accident Risk Association. .	120
4.71	ROC-AUC curve - Random Forest Classifier (Trento)	123
4.72	SHAP output for RFC model (Trento)	123
4.73	Zones in Trento categorized by high or low accident risk.	124
4.74	ROC-AUC curve - Random Forest Classifier (Bolzano)	125
4.75	SHAP output for RFC model (Bolzano)	125
4.76	Zones in Bolzano categorized by high or low accident risk.	126

LIST OF FIGURES

List of Tables

3.1	Summary of OSM Data Downloaded for the different analysis	30
3.2	Description of LTS Rules	43
3.3	Tags used for POI extraction	56
3.4	Categories, Types, and Weights for Destination Scoring	59
3.5	Scoring Processes for Low Stress Destinations	60
4.1	Summary of Exploratory Data Analysis Metrics for Trento	70
4.2	Summary of Exploratory Data Analysis Metrics for Bolzano	71
4.3	Benefit Metric for Identified Gaps	102
4.4	Benefit Metric for Identified Gaps	104
4.5	Summary of components by quadrant in Trento	117
4.6	Summary of components by quadrant in Bolzano	119
4.7	Classification report for Logistic Regression	121
4.8	Classification report for Random Forest Classifier	121
4.9	Classification report for Support Vector Machine	121
4.10	Classification report for Random Forest Classifier - Full Dataset	122
4.11	Classification report for Random Forest Classifier - Full Dataset	125

