Lisbon School of Economics & Management
Universidade de Lisboa

MASTER

DATA ANALYTICS FOR BUSINESS

MASTER'S FINAL WORK

DISSERTATION

Resilience of the Metro of Porto and Accessibility of POIs: A Network Science Approach

Teresa Maria Nugent Ribeiro Matos Gomes

February 2025

MASTER
DATA ANALYTICS FOR BUSINESS

MASTER'S FINAL WORK
DISSERTATION

Resilience of the Metro of Porto and Accessibility of POIs: A Network Science Approach

Teresa Maria Nugent Ribeiro Matos Gomes

SUPERVISOR: Prof. Carlos J. Costa
CO-SUPERVISOR: João Tiago Aparício

JURY:

PRESIDENT - Prof. Mário Caldeira
MEMBER - Prof. Flávio Costa Romão
SUPERVISOR - Prof. Carlos J. Costa

February 2025

Glossary

CRISP-DM – Cross-Industry Standard Process for Data Mining.

CRS – Coordinate Reference System

GTFS – General Transit Feed Specification.

LCC – Largest Connected Component

MFW – Master's Final Work

OSM – Open Street Map

POIs – Points of Interest.

ABSTRACT

Over the past decades, tourism in Portugal has grown exponentially, particularly in major cities such as Porto. As a result, ensuring smooth mobility for tourists and providing them with efficient access to the city's main points of interest is important. This issue not only affects tourists but also directly impacts on the daily lives of Porto residents. To this end, this Master's Final Work goal is to address two research questions: "How resilient is Porto's Metro system?" and "To what extent are Points of Interest (POIs) in Porto accessible via the Metro?".

To assess the Metro's resilience, simulations were conducted by removing stops based on betweenness centrality, degree centrality, and random removal strategies, in addition to considering percolation thresholds. These simulations were carried out using network science techniques, specifically analyzing the largest connected component of the Metro network. To evaluate the accessibility of POIs via the Metro, the walking distance between Metro stops and the POIs was calculated. This analysis was based on a 15-minute walking threshold to determine the Metro's coverage of major attractions.

The findings related to the first research question show that Porto's Metro system lacks resilience to targeted disruptions at stops with high betweenness and degree centrality, leading to an almost immediate impact on the connectivity of the Metro lines. In contrast, randomly selected stop removals exhibit a more stable pattern, causing a more gradual decline in network connectivity. Regarding the second research question, the results reveal that Metro stops cover at least 71.75% of the city's main points of interest under the applied thresholds.

These insights could influence transportation planning in Porto and may also be extended to other cities significantly affected by increasing tourism.

KEYWORDS: POIs; Accessibility; Resilience; Network.

JEL CODES: C63; C65; D85.

RESUMO

Ao longo das últimas décadas, o turismo em Portugal tem crescido exponencialmente, especialmente nas grandes cidades, como é o caso do Porto. Como tal, garantir que a mobilidade dos turistas ocorre de forma fluida e que tenham acesso aos principais pontos turísticos da cidade é de extrema relevância. Afeta não apenas os turistas, mas também os habitantes da cidade. Com esse intuito, foi posto em prática, ao longo deste Trabalho Final de Mestrado a resposta a duas questões: "O quão resiliente é o Metro do Porto?" e "Até que ponto os principais pontos de interesse no Porto são acessíveis através do Metro?".

Para responder à questão da resiliência foram simuladas remoções de estações do Metro do Porto tendo em conta as estratégias de centralidade de intermediação, centralidade de grau e remoção aleatória, e adicionalmente foi também considerado limiares de percolação. Para todas estas estratégias foi tido em conta técnicas de ciência de redes e o maior componente ligado. Para responder à questão da acessibilidade aos pontos de interesse através do Metro foi considerada a distância de caminhar entre o Metro e os pontos de interesse. Esta distância teve em conta o limar de 15 minutos a pé para avaliar a cobertura dos POIs.

Os resultados da primeira questão revelam que o Metro do Porto não demonstra resiliência a disrupções direcionadas às estações de maior centralidade de intermediação e de grau resultando num efeito quase imediato sobre a conectividade das linhas. Em contrapartida as disrupções em estações selecionadas aleatoriamente apresentam um comportamento mais estável, provocando um impacto mais gradual na conectividade do Metro. Os resultados da segunda questão indicam que, considerando os limiares aplicados, pelo menos 71,75% dos pontos de interesse estão abrangidos pelas estações do Metro do Porto.

Estes resultados podem influenciar o planeamento dos transportes na cidade do Porto e até ser aplicados a outras cidades fortemente impactadas pelo crescimento do turismo.

PALAVRAS CHAVE: POIs; Accessibility; Resilience; Network.

JEL CODES: C63; C65; D85.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Firstly, I would like to say thank you to Professor Carlos J. Costa and João Tiago Aparício for all the valuable guidance, help, and insights that they provided throughout this Master's Final Work.

Secondly, I want to express my gratitude to the women who made all this possible. To my mother, that did all and above for me to be able to enroll in this important, personally and professionally, chapter of my life, and for all the unconditionally love that she gave me throughout my life. To my grandmother, for welcoming me into her home and making everything feel easy. To the friendships that ISEG gave me, Ana and Marta, who have turned these past two years into one of the most memorable and meaningful experiences of my life. To my sister, for always being my constant source of encouragement, compassion, and belief in my dreams.

Additionally, I want to say thank you to my partner, Diego, and to all my family and friends that were patient and supportive during all this process.

Lastly, I want to acknowledge my father, who left us far too soon, without being able to see me graduate and grow as a woman. Though he is no longer here, I find comfort in knowing that he was proud of me.

A final shoutout to generative AI tools for the help provided to refine the language used in the MFW. While these tools assisted in refining the language, all intellectual content and reasoning remain my own.

1.INTRODUCTION

1.1 Context and Motivation

Over the past few decades, tourism in Portugal has experienced significant growth, particularly in the country's second-largest city, Porto. Whether it is sunny weather, world-renowned wine, stunning landscapes, affordable yet exceptional restaurants, or warm and welcoming people, something about Porto continues to captivate visitors from around the world. While it is not fully understandable what drives everyone's ambition to visit this vibrant city, several factors like this likely play a role. On the one hand, the increase in tourism could be attributed to global trends, such as greater accessibility to travel, the opening of international borders, and a rise in disposable income, allowing more people to explore the world. On the other hand, a deeper, more emotional factor is also at play - especially in the wake of the COVID-19 pandemic. The pandemic reshaped priorities for many, leading to a collective realization of the importance of enjoying life's experiences, such as traveling and discovering new cultures. Whatever the reasons may be, one thing is clear: tourists are choosing more often Porto as a vacation destination. For these reasons, along with the daily life of Porto's residents—workers, students, and others—and my personal perspective as a citizen of Porto and a frequent user of its Metro system and POIs, studying the resilience of the Metro and the accessibility that promotes to the POIs is particularly meaningful.

A resilient Metro system is vital because it serves as a primary mode of transportation for both locals and tourists navigating the city. It is essential that the Metro is prepared to handle disruptions and maintain functionality daily. In connection with this, the accessibility of POIs through the Metro network is equally important. POIs should be within reasonable walking distance from Metro stops, ensuring that reaching these locations is convenient and not overly challenging.

POIs are what attract tourists to a city—they could be a statue, a museum, a place of worship, or even a wine cellar in Porto. Ensuring good accessibility to these POIs improves visitors' experience and benefits residents, making it easier for everyone to engage in and enjoy what the city has to offer.

Having both a resilient Metro system and easily accessible POIs via the Metro is essential for urban planning. It contributes to a well-connected city, promotes environmental and health benefits by encouraging the use of public transportation and walkability, and adds to Porto's overall convenience and livability for its residents and visitors.

1.2 Research Questions

This MFW aims to explore and address these two critical research questions:

- How resilient is Porto's Metro system?
- To what extent are Points of Interest (POIs) in Porto accessible via the Metro?

These questions aim to evaluate the Metro's ability to handle disruptions and its effectiveness in connecting people to the city's main attractions. Understanding these aspects will provide valuable insights into the reliability and accessibility of Porto's urban infrastructure.

To address these questions, the study will use several analytical techniques. To calculate the resilience of Porto's Metro system, the impact of node (stop) removals will be examined through three strategies: random removal, removal based on betweenness centrality, and removal based on degree centrality. Additionally, percolation analysis will be applied to evaluate network fragmentation and critical failure points further. For the second research question, which focuses on accessibility, the analysis will incorporate the concept of 15-minute walkability, assessing how well the Metro system connects people to key locations within a reasonable walking distance.

1.3 Structure of the Master's Final Work

This MFW is organized into several key phases to answer the research questions. It begins with a comprehensive literature review, exploring existing knowledge on network science, transportation resilience, Porto's Metro system structure, accessibility, and POIs. Next comes the methodology section that explains the steps considered in this MFW, following the CRISP-DM framework, which includes business understanding, data understanding, data preparation, modeling, and evaluation. Following this, the results and discussion section presents the findings and their implications, leading to the final phase, the conclusion, which also addresses the study's limitations and potential future research.

2. LITERATURE REVIEW

2.1 Network Science Fundamentals

Network Science can be used for multiple purposes. From the study of neurology and epidemic spread to computer science, social interactions, transportation behaviors, and analyzing population density in a country. Because of its wide range of applications, its definition is not specific but rather one that depends on the study in matter. According to the National Research Council (cited in (G. Lewis, 2011)) it can be described as an "organized knowledge of networks based on their study using the scientific method".

In simpler terms, Network Science is a set of objects that are connected to each other that all together represent something abstract but real, like social connections, the spread of diseases or transportation routes. These connections are visualized as graphs, which is where graph theory comes into play. (G. Lewis, 2011)

Graph theory is basically a way to describe the interactions and the structure between nodes with the use of a graph like in Figure 1. This graph is composed of nodes, edges, and the mapping function. The nodes, also referred to as vertices or actors, can represent, for example, a human being, a location or a molecule. On the other hand, the edges, also known as arcs, links, or relationships, can represent connections between humans, roads between locations, or interactions between molecules. (G. Lewis, 2011)

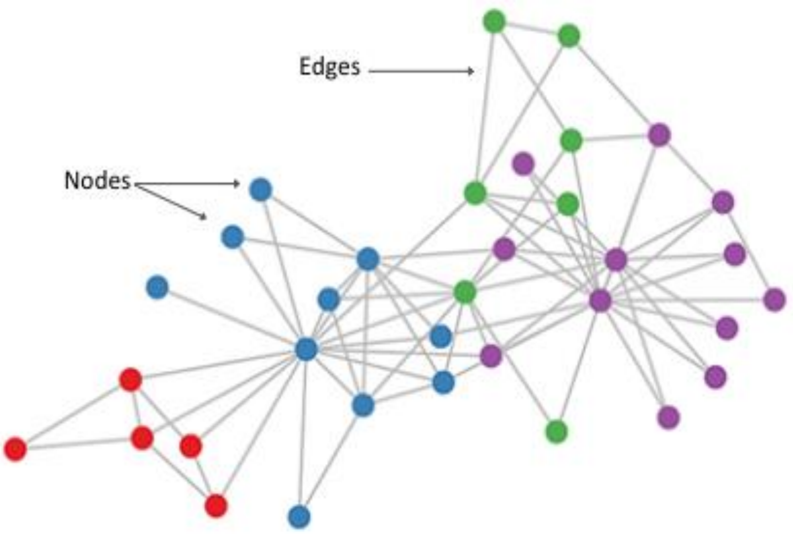

FIGURE 1- Network Graph

The purpose of the mapping function is to describe the structure of the graph, showing how each node is linked to other nodes through connections. The graphs can be either static or dynamic. In dynamic graphs, the components (nodes, edges, and the mapping function) change over time, meaning the network changes due to its dynamic nature rather than its original structure. In contrast, static graphs have components that remain unchanged over time. (G. Lewis, 2011)

Altogether, the purpose of the graph is to illustrate a particular behavior or pattern. In mathematical or programming terms, these graphs are structured as follows:

$$(1)\ G(t) = \{N(t), L(t), F(t) : J(t)\},$$

Where variables in Equation (1) are G representing the network graph, t is the time associated with the graph, N is the node, L is the edge, and $F = N \times N$ represents the mapping function. (G. Lewis, 2011)

However, graph theory alone is not sufficient to describe a Network. In addition to the graph's structure (nodes, edges, and the mapping function), dynamic properties are also needed. These are composed of microrules and macrorules. The micro rules help to describe local behaviors within the network, such as interactions between specific nodes. The macro rule helps to describe global patterns found in the network that emerge from these local interactions, giving a picture of the overall system. (G. Lewis, 2011)

2.2 Network Properties

To better understand the behaviors and characteristics of the network studied in this MFW, certain properties and concepts must be considered. These properties may relate to the network as an overall system, to the graph structure, or specifically to the nodes.

The graph within a network can be either directed or undirected. In a directed graph, each edge has a specific direction that must be followed, creating a distinct path from one node to another; therefore, the order of the nodes is important. In this case, each edge has a defined head and tail. In contrast, an undirected graph has no directional edges, so the order of the nodes is unimportant, allowing free movement between connected nodes. (G. Lewis, 2011)

Nodes in a graph provide significant insights into the network's topology and characteristics. The influence or importance of a specific node can be understood from its degree, which is the number of edges connected to that node. The more edges a node has, the higher its degree. A node with the highest degree is referred to as a hub. (G. Lewis, 2011)

Beyond node degree, concepts like path, circuit, loop, degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality also play an essential role in understanding a network's structure.

The path represents the route of edges that connect one node to another in a network, with its length, also referred to as a hop, being equal to the number of edges between them. Normally, shorter path lengths indicate more closely connected nodes, and the longest path in the graph is called the diameter of the graph. The paths are important as they provide insights into the network connectivity and help identify crucial nodes that act as bridges between distinct elements within the network. (G. Lewis, 2011)

The circuit and loop are similar concepts, both involving paths that start and finish in the same node. However, circuits require a path length of at least two, while the loop has a path length of one, meaning a self-connecting node. (G. Lewis, 2011)

The degree centrality of a node equals the sum of all direct edges it has. The higher the degree centrality, the more connected and influential the node is within the network. (G. Lewis, 2011)

Closeness centrality measures how near a node is to all other nodes in the network. The lower a node's closeness centrality score, the more central it is, allowing it to communicate more efficiently within the network without passing through many intermediate nodes. In contrast, the higher a node's closeness centrality score, the less central and communicative it is with other nodes. (Derr, 2021)

Betweenness centrality measures how a node serves as a connecting bridge to other nodes. The higher a node's betweenness centrality score, the more central it is. In this case, the paths to get to other nodes are shorter, which allows the node to hold influence and control within the network. In contrast, a lower betweenness centrality score implies that a node is less central, with longer paths to reach other nodes. It means that it plays a less significant role in connecting different parts of the network and holds less influence and importance. (Derr, 2021)

Lastly, eigenvector centrality is distinct from closeness and betweenness centrality because it does not assess a node's centrality within the network itself. Instead, it measures the proximity to nodes that are influential in the network. A higher eigenvector centrality score suggests that a node is closer to significant and important nodes within the network, although it may not be central itself. (Ruhnau, 2000)

The power of a network is measured not by the degree of individual nodes but instead by the strength of the overall set of edges and nodes. In many cases, these sets of edges and nodes create modules inside the network. To evaluate the power of these modules, it is also important to consider the following concepts: modularity, assortative and disassortative networks.

Modularity measures the strength of connections within modules in a network. These modules, also known as communities or clusters, are crucial since they can reveal how different parts of the network interact and influence each other. These modules are also important to test the network resilience, as they show how the network might respond in case of unusual activity within the nodes or the impact when one or multiple nodes are removed. The higher the modularity, the more connected the nodes are within each module and the less connected they are to nodes from different modules. When a network has high modularity, it indicates a well-defined community structure, where nodes are grouped into distinct, cohesive clusters with relatively few edges connecting these groups. (Dinh et al., 2015) (Buldyrev et al., 2010)

The assortative networks have well-defined clusters, where nodes within each cluster tend to connect to other nodes with similar characteristics. In these networks, high-degree nodes typically connect with other high-degree nodes, and low-degree nodes connect with low-degree nodes. Such networks are commonly found in social contexts, like friendship and professional relationships. (Moore et al., 2011)

In contrast to the assortative networks, when nodes within clusters tend to connect with others that contain different characteristics, it is considered a disassortative network. High-degree nodes typically connect with low-degree nodes. Such networks are commonly found in transportation routes or technological services. Despite their different characteristics, both types of networks can coexist within the same system and provide insights that help categorize the network's overall strength. (Moore et al., 2011)

2.3 Network Science Applicability

While all the characteristics of the network science approach are important, it is equally crucial to understand why this approach is particularly well-suited for evaluating the resilience of Metro systems.

The structure of a Metro system can be directly compared to a network, as it consists of stops (nodes) connected by lines (edges). This similarity allows the Metro network to be modeled as a graph with interconnected nodes and edges. Such a representation enables the analysis of the system's topology, the resilience of individual stops, its overall efficiency, the identification of critical stops or routes, and the potential impact of adding or removing new stops on the Metro performance.

By analyzing various cases, such as the Shanghai Metro study (Feng et al., 2024) and the Zhengzhou Metro study (Qi et al., 2022), it becomes clear that a well-structured transportation network is essential for ensuring the smooth circulation of users and for being prepared to handle disruptions within the network.

These studies demonstrate the power of network science in evaluating Metro system resilience under different disruption scenarios. For example, the Shanghai Metro study examined the relationship between waterlogging, the increase in rainfall, and the number of affected stops, concluding that extreme rainfall significantly reduces network performance. Similarly, the Zhengzhou Metro study analyzed the robustness and vulnerability of the network under both targeted and random node failures, revealing that central nodes and interchange stops are crucial for maintaining connectivity.

Both these cases emphasize the importance of designing transportation systems that not only support efficient daily operations but also ensure resilience in the face of unexpected circumstances.

A well-designed system is fundamental for societal development, economic growth and public safety, as it serves as a means of transportation for work, schools, and other points of interest around the cities. (Wassmer et al., 2024) For that reason, understanding the resilience of these networks is crucial for ensuring their sustainability and continuous operation, especially in the face of disruptions.

2.4 Resilience in Transport Networks

2.4.1 Definitions

Resilience, or in simpler words, flexibility, is a crucial concept for this MFW, as it will help to understand whether the network is resilient or not, as well as both external and internal factors that could disrupt its functionality. Therefore, understanding its definition and the factors that can affect resilience is essential for the analysis.

Resilience can have various interpretations depending on the field or situation in question. In a broader context, it can be described as the capability to overcome the challenges and adversities that individuals or systems face daily. (Amghar et al., 2023) In contrast, when applied to the transportation sector, according to various authors such as Pan et al. (cited in Amghar et al., 2023) and Imran et al. (cited in Amghar et al., 2023), resilience is defined as the ability of a transportation system to incorporate unexpected interferences, while simultaneously resisting and adapting to these challenges, in order to either recover or to maintain the demand of the service.

2.4.2 Factors That can Affect the Resilience of a Transport System

Since transport systems are composed of multiple lines and stops, it is natural for them to be exposed to factors that can compromise their normal functioning. Such factors can be both external and internal. External action can be done either on purpose, such as in wars or terrorism, or unintentionally, such as in natural disasters (e.g., floods). The internal, on the other hand, may include operational errors, system failures, or strikes by employees. Each factor has a distinct impact on the transport system and requires a personalized approach. However, they share a common element: all these factors influence the daily experience of transportation users and, on a broader scale, can lead to substantial financial consequences. This emphasizes the need to assess the resilience of transport systems against factors that may disrupt their functionality. (Amghar et al., 2023)

The concept of resilience has been explored in various projects, such as the Shanghai Metro study (Feng et al., 2024) and the Zhengzhou Metro study (Qi et al., 2022). These studies provide valuable insights into how external and internal disruptions impact the resilience of the Metro systems. Both cases demonstrate the importance of analyzing the structure and robustness of Metro networks to ensure their reliability during disruptions. Inspired by these conclusions, applying a similar approach to the Porto Metro system would allow for a deeper understanding of its resilience and determine how well it can handle potential challenges.

2.4.3 Techniques to Evaluate Resilience

When evaluating the resilience of transport networks, it is essential to assess how disruptions impact overall connectivity. A widely used metric in network vulnerability analysis is the Largest Connected Component (LCC), which represents the largest remaining part of the network that maintains its connectivity after a disruption. The LCC size provides insight into the system's ability to remain functional when nodes or edges are removed. A decrease in the LCC size suggests that the Metro network is becoming more fragmented, making it less efficient and harder for passengers to move through the system. One of the most effective ways to study network robustness is through targeted and random attack simulations. These approaches simulate real-world disruptions to understand how the system responds to different threats. According to Lekha and Balakrishnan (2018), network attacks can be broadly categorized into two types. The first type, random attacks, removes nodes unpredictably, mimicking natural failures such as power outages or isolated service disruptions. The second type, targeted attacks, systematically eliminates the most critical nodes based on predefined criteria such as degree centrality or betweenness centrality, replicating the impact of deliberate sabotage or large-scale network failures.(Lekha & Balakrishnan, 2018)

A key idea in resilience analysis is the percolation threshold, which marks the point where a network starts breaking apart. Percolation theory helps study how removing stops (nodes), or connections (edges) affects a network's structure. As more nodes or edges are removed, the network gradually shifts from fully connected to a scattered system with isolated parts. In transport networks, percolation analysis is particularly relevant for understanding network resilience and identifying the point at which the service reliability becomes compromised. In large-scale transportation networks, the percolation threshold indicates the fraction of stops or connections that must be removed before the Metro system experiences a major breakdown in connectivity. (Rong et al., 2022)

The findings in Lekha and Balakrishnan (2018) highlight that removing nodes based on betweenness centrality and degree centrality has the most significant impact on the LCC size. For this reason, these two strategies will be used in this study. By applying these concepts to the Porto Metro system, percolation analysis and LCC tracking provide valuable insights into the network's resilience under different failures. These insights are particularly important for transport planners, as they can guide the development of improvements in the Metro systems.

2.5 Structure and Topology of the Porto Metro Network

The Metro of Porto, unlike many major cities around the globe, has most of its line above ground. This characteristic makes the network more susceptible to accidents and malfunctions, which can compromise the safety of the passengers and people in the surroundings. (Metro do Porto, 2024a)

The Metro network is composed of six lines (A, B, Bexp, C, D, E and F) and eighty-five stops that cover seven municipalities: Gondomar, Porto, Maia, Matosinhos, Valongo, Vila do Conde e Vila Nova de Gaia. The entire Metro network has an extension of approximately 70km, where 8.5km are underground. According to the official Metro of Porto website, each line is capable and designed to accommodate up to 90 000 passengers per hour. In 2023, approximately 79 million validations were recorded, resulting in an average of 219 444 passengers daily. Both these factors highlight the importance of maintaining a resilient network capable of handling such a high volume of users. (Metro do Porto, 2024a)

The topology of the Metro network may not stand out in terms of dispersion and coverage, as approximately thirteen stops are common across five of the lines, from "Senhora da Hora" to "Campanhã". However, this does not take away from the importance of studying it since removing any of these thirteen stops during peak hours due to malfunctions or other factors would significantly impact the daily flow of passengers. (Metro do Porto, 2024a)

2.6 Accessibility and Points of Interest

2.6.1 Accessibility

Accessibility can be defined in various ways depending on the context, but in this case, the focus is on accessibility within an urban context. According to Guida (2020), urban accessibility incorporates not only the ease of reaching a particular destination but also the availability and quality of different methods to get there. These methods can include well-designed pedestrian pathways, efficient public transportation systems such as Metro and bus networks, or well-maintained road infrastructure.

The concept of the "15-minute city", as explored by Aparicio et al.(2024), is a valuable framework for understanding accessibility to POIs in Porto from Metro stops. This concept accentuates the importance of having essential services and amenities within a 15-minute walking distance. While the original study applies this idea to broader urban planning, it can also be adapted here to assess Metro-based accessibility. By evaluating whether POIs near Metro stops fall within a walkable radius, this principle helps determine how accessible these locations truly are.

To calculate the distance covered in a 15-minute walk, it was considered the average fastest walking speed. According to Cronkleton, (2019), the fastest speed was estimated at 1.43 meters per second. Over 15 minutes, this walking speed corresponds up to 1 287 meters.

2.6.2 Points of Interest

Point of interest (POI) is a location that holds significance for individuals, groups, or communities, often due to its cultural, historical, or functional importance. These locations can include various types of places such as restaurants, cafés and other social spaces. POIs also include commercial sites like stores and markets, essential public institutions such as hospitals, schools, and government offices, as well as historical landmarks like monuments and museums. Additionally, transportation hubs, including airports and train stops, are vital POIs that facilitate travel and mobility. (Rafaqat, 2023)

POIs are important not just for community life and economy but also for tourism. They help shape the overall experience for both visitors and residents by creating spaces for social interaction, commerce, and culture. Beyond attracting tourists, these places play a central role in defining the identity of a region and supporting its infrastructure. From historical landmarks to transport hubs, POIs contribute to the social life and economic well-being of a city, making them crucial for both everyday life and tourism.

2.6.3 How do Transport Networks Facilitate Access to POIs

Why is good accessibility so important for Porto tourism? In recent years, Porto has experienced significant growth in its tourism sector, consistently earning recognition as one of Europe's top vacation destinations. This increase in tourism has not only elevated the city's global profile but also contributed substantially to its economy, driving increased revenue through hospitality, transportation, and cultural activities. This highlights the importance of a well-developed transport infrastructure, which is essential not only for accommodating the city's population but also for addressing the increasing needs of tourists.

In a study made by Tan and Ismail (2020), they explored the correlation between transport and tourism, emphasizing why strong transport systems are so important. While transportation can function on its own without tourism, tourism cannot succeed without reliable travel options. Transportation directly impacts on tourist satisfaction, which is critical for a destination's success. The more convenient it is to reach POIs, the more satisfied visitors tend to be. This makes transportation a key factor in creating positive travel experiences and effectively marketing a destination.

3. METHODOLOGY

The aim of this chapter is to outline the methodology adopted to address the proposed research questions: How resilient is Porto's Metro system, and how accessible are the Points of Interest (POIs)?

As with any project, a data science project also requires a structured and organized approach. Therefore, the CRISP-DM methodology was adopted as a base for developing and implementing this project. CRISP-DM, which stands for Cross-Industry Standard Process for Data Mining, is a widely used framework for managing the lifecycle of data science projects (Costa and Aparicio, 2020, 2021). The methodology comprises six phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment, as illustrated in Figure 2. These phases are not strictly linear, as real-world projects often encounter challenges that require revisiting earlier stages. For instance, during the evaluation phase, it may be necessary to revisit the business understanding to reassess the project's objectives or refine the desired outcomes. This iterative approach ensures optimal performance during the deployment phase. (IBM, 2021) (Costa & Aparicio, 2020)

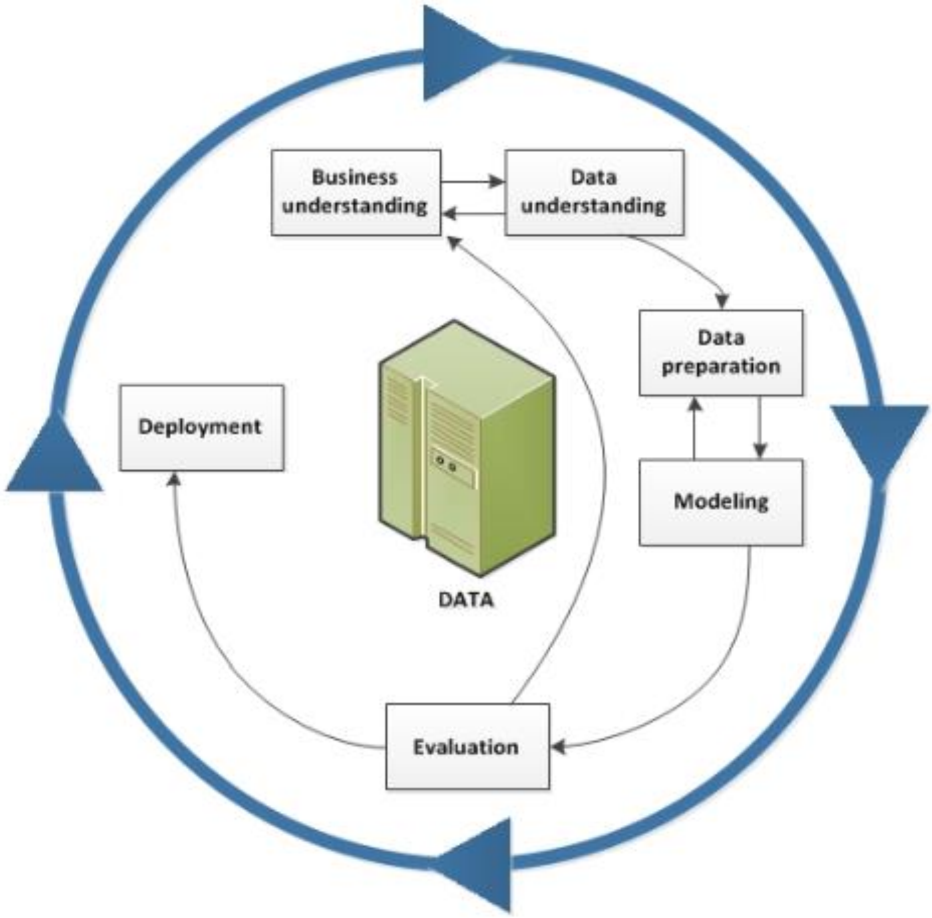

FIGURE 2 – Cross - Industry Standard Process for Data Mining (IBM, 2021)

3.1 Business Understanding

The primary focus of this phase is to understand the research questions and their significance within the context of this MFW. (Bemthuis et al., 2024)

The first research question, "How resilient is Porto's Metro system?" as referred previously in the introduction chapter, investigates the ability of the Metro system to withstand disruptions and maintain service continuity. Given the critical role of the Metro in Porto on the daily mobility of both residents and tourists, understanding its resilience is essential for ensuring the smooth and uninterrupted flow of people within the city.

The second research question, "To what extent are Points of Interest (POIs) in Porto accessible via the Metro?" explores the accessibility of essential services, amenities, and attractions surrounding the Metro network. Addressing this question is crucial for several reasons. Firstly, accessible POIs minimize the reliance on private vehicles. Secondly, it ensures that essential services, amenities, and cultural attractions are easily reachable from the Metro network, maintaining a well-connected and vibrant city.

By addressing these research questions, this MFW aims to provide valuable insights into the effectiveness and accessibility of the Porto Metro system, contributing to a better understanding of urban mobility and urban planning within the city.

3.2 Data Understanding

In this phase, the primary objective is to collect the data and the necessary tools to address the research questions. (Bemthuis et al., 2024) Two types of data were used for this purpose.

To address the first research question, "How resilient is Porto's Metro system?", the foundational dataset was obtained from the Metro of Porto data portal, which provides extensive information about the schedules, stops, and routes. Specifically, this project used a GTFS file from this portal, updated on September 6, 2024. (Metro do Porto, 2024b)

GTFS, which stands for General Transit Feed Specification, is a standardized data format that enables transportation agencies to describe the components of their services. (Delgado Rodriguez, 2024) The GTFS file for the Metro of Porto includes the following datasets in a text document format: Agency, Calendar, Calendar Dates, Fare Attributes, Fare Rules, Routes, Shapes, Stop Times, Stops, Transfers and Trips. To read, process, and analyze all these files the chosen tool was Python in a Jupyter Notebook environment.

However, not all these datasets were used in the analysis, as certain elements, such as Agency, Calendar, Calendar Dates, Fare Attributes, Fare Rules, and Transfers, were considered non-critical for addressing the research question.

For the datasets deemed critical, the following were used in the analysis. The Routes dataset contains information about all the routes within the Metro system, identified by route IDs such as A, B, Bexp, C, D, E, and F. The Shapes dataset provides details about the precise geographic paths of the Metro lines, represented using longitude and latitude coordinates. The Stop Times dataset contains information that serves as a bridge between the Stops and Trips datasets by connecting them through stop_id and trip_id attributes. The Stops dataset includes comprehensive information about all Metro stops in Porto, including the stop ID, stop name, zone ID, and the respective longitude and latitude coordinates. Lastly, the Trips dataset connects various elements using the route ID and provides additional attributes such as the service ID and trip head sign, which specifies the destination of each trip.

To provide a clearer visualization of the connectivity between the datasets and the attributes used to establish these connections, the following map was created:

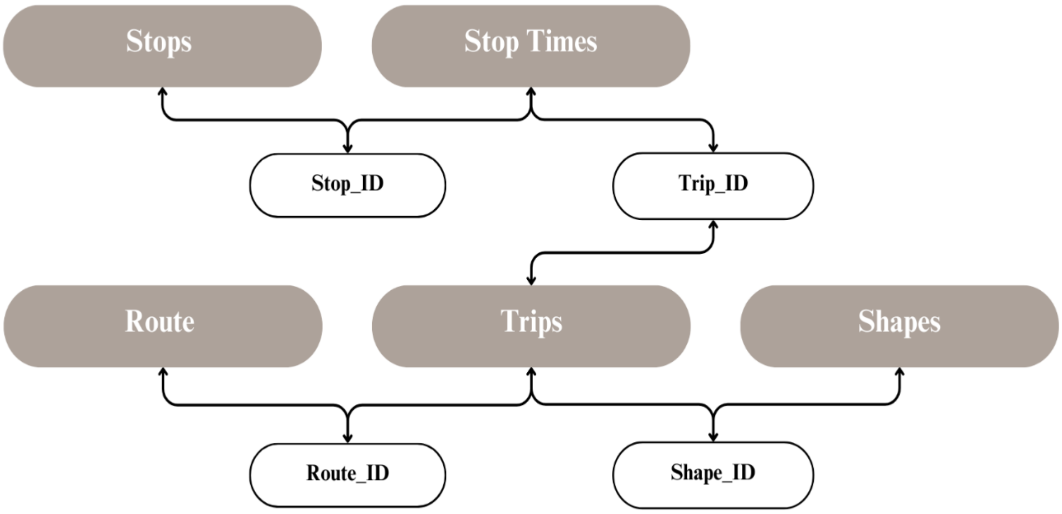

FIGURE 3 - Connections Between GTFS Datasets

Caption:

- Colored figures – datasets from the GTFS that were used in the analysis.
- White figures – attributes that connect each dataset.

To address the second research question, "To what extent are Points of Interest (POIs) in Porto accessible via the Metro?", it was used the dataset extracted from the Overpass Turbo. Overpass Turbo is a tool that allows users to extract data from the OpenStreetMap (OSM) database. It uses the Overpass API, which enables data to be filtered based on geospatial coordinates, including latitude and longitude.

The results appear in two formats, as a dictionary with the POIs information (in {}, which represents the data in a structured format, such as JSON or XML) and as a map, as shown in Figure 4, allowing for both numerical and visual representations of the POIs. (OverPassTurbo, 2024)

To gather the data needed to answer this research question, the code below was run in Overpass Turbo:

```
[out:json][timeout:25];
(
  node["tourism"](around:40000,41.14961,-8.61099);
  way["tourism"](around:40000,41.14961,-8.61099);
  relation["tourism"](around:40000,41.14961,-8.61099);
);
out body;
```

FIGURE 4 - Code to Extract Data from OverPassTurbo

This code extracts data in JSON format, specifically for elements where the node, way, and relation are related to tourism. The query "*around: 40000, 41.14961, -8.61099*" retrieves tourism-related elements within a 40 km radius of the coordinates for Porto (41.14961, -8.61099). These coordinates for Porto are according to Latitude.to (2024). The query "*out body*" retrieves POIs with detailed information, including attributes such as the name, ID, type, coordinates, website, opening hours, and more. When executing the code, the map format result is illustrated in Figure 5.

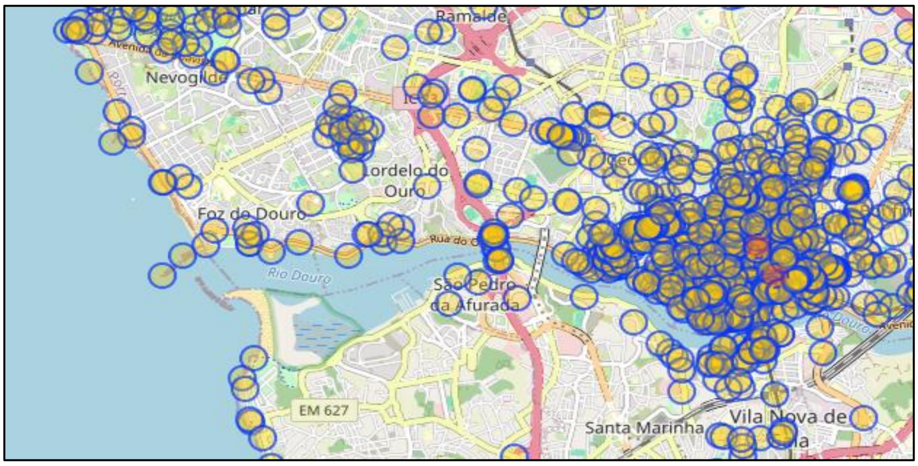

FIGURE 5 – Result in a Map Format from OverPassTurbo

When executing the Overpass Turbo query, the results are in a dictionary format and contain detailed information about the POIs. This dataset contains attributes, including information such as the unique ID, name, address, type of amenity, website, contact details, and more. However, for the purpose of this analysis, only the attributes id, addr: city, amenity, name, tourism, and geometry are considered relevant.

The id attribute represents the unique identifier assigned to each POI, allowing for precise referencing of individual entries. The addr: city attribute indicates the city where the POI is located, ensuring that only POIs within Porto's boundaries are included in the analysis. The amenity attribute describes the type of amenity the POI represents, which, in this dataset, includes categories such as theatre, restaurant, fountain, place of worship, bar, clock, bench, or "None" if the POI does not belong to a specific amenity category. The name attribute provides the name of the POI, while the tourism attribute identifies the type of tourism-related feature the POI represents, such as a hotel, artwork, viewpoint, or other similar categories. The geometry attribute contains spatial information that defines the exact location of each POI. This spatial data is represented in two possible formats: the point format and the Line String format. The point format is used for POIs that are represented by a single location, such as monuments or restaurants, and includes the latitude and longitude coordinates of the POI. The Line String format, on the other hand, is used for POIs that represent linear features, such as trails or extended areas and includes a series of latitude and longitude coordinates that define the shape of the POI.

3.3 Data Preparation

In this phase, as the name suggests, the goal is to prepare the data by performing tasks such as cleaning and organizing it, to ensure it is in the most suitable format for subsequent modeling and evaluation. (Bemthuis et al., 2024)

3.3.1 GTFS Files

The first step in preparing the data from the GTFS files involved uploading them into Jupyter Notebook. Once uploaded, the files were read to explore their contents. During this process, it was observed that some tables contained missing values, which could impact the analysis. As a result, a data cleaning step was performed to remove unnecessary columns and those with missing values. This cleaning process was applied to the routes, shapes, stop_times, stops, and trips tables, as these were the only tables required for the analysis.

To better understand the data, a code was executed to identify the unique values within these five tables, enabling an understanding of how the tables are connected. Such connections can be observed in Figure 3. Next, the frequency of stops appearing in the stop_times table was calculated to identify potential hubs within the Metro network. The results showed that stop Trindade and Senhora da hora are a hub or nodes with high connectivity, making it critical points in the Metro system. Following this, the number of trips per route was analyzed to understand how trips are distributed across different Metro routes. The analysis revealed that Route D, which connects Hospital de S. João to Laborim, is the busiest route, with the highest trip count (993 trips). This suggests that Route D plays a key role in the Metro network. On the other hand, Route Bexp, which connects Póvoa de Varzim to Estádio do Dragão in less time than Route B, was identified as the least busy route, with only 54 trips. This is likely because Route Bexp operates less frequently, with trips scheduled every 30 minutes. To facilitate further analysis, the stops, stop_times, trips, and routes tables were merged into a single dataset. This consolidation made the data more accessible and easier to work with for subsequent steps. Finally, a network graph was created using the Matplotlib library to visualize the Metro system. In this graph, nodes represent Metro stops, while edges represent their connections, as shown in Figure 6.

This visualization provides a clearer understanding of the Metro network's structure.

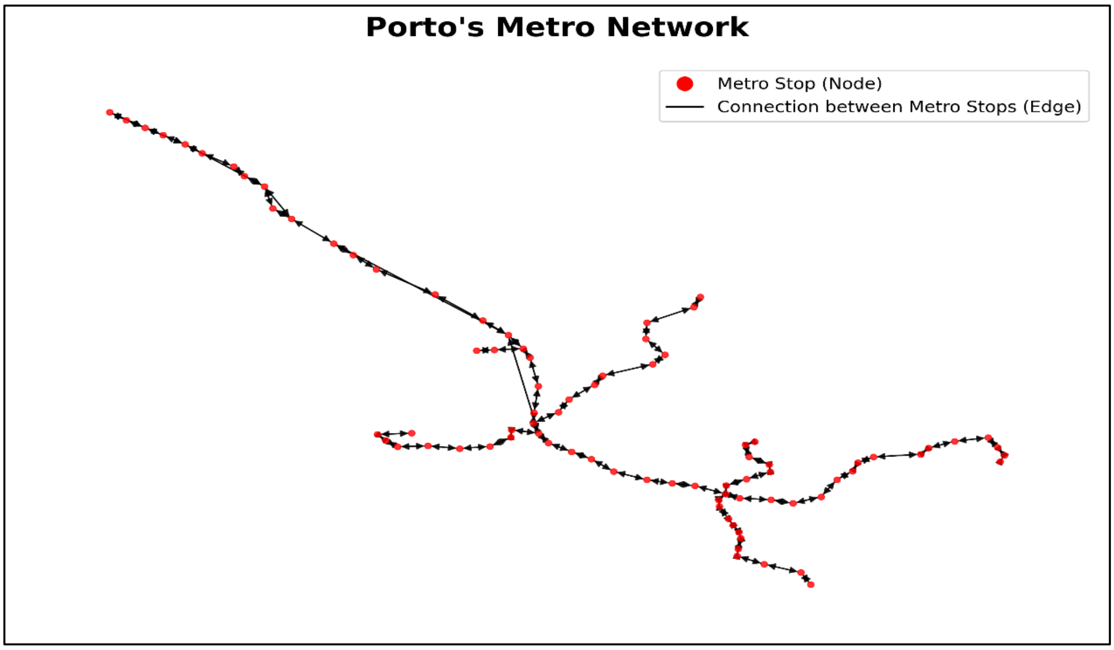

FIGURE 6 – Porto's Metro Network

To better understand the structure and components of the Metro network, several metrics and analyses were performed. These include calculations of node degree, determining whether the graph is directed or undirected, measuring path lengths, degree centrality, closeness centrality, betweenness centrality, the identification of the existence of clusters, and whether the graph is assortative or disassortative.

For the node degree, calculations revealed that the maximum degree in the network is 8, meaning some nodes have 8 connections (both incoming and outgoing). An example is stop Trindade, previously identified as a hub, where multiple Metro lines intersect. On the other hand, the minimum degree in the network is 2, observed at stop Fânzeres, which is located at the end of a line and has one incoming and one outgoing connection.

The analysis also confirmed that the graph is directed, which accurately reflects the real Metro system, as each direction has its own line rather than a shared bidirectional line. This distinction means that disruptions in one direction of a stop do not necessarily impact the opposite direction in the same way.

Regarding path lengths, the average shortest path length in the network is approximately 13 stops. This indicates that, on average, passengers traveling via the shortest route will pass through about 13 stops. In contrast, the diameter of the network (the longest shortest path) is 34 stops, meaning the two most distant stops in the network require 34 stops to connect via the shortest route.

When calculating degree centrality, the maximum value is 0.0952, observed at stops such as Trindade and Senhora da Hora. This indicates that these stops are the most influential or well-connected nodes in the network, playing a critical role in the system. In contrast, the minimum degree centrality value is 0.0238, seen at stops like Hospital São João, which is an end-of-line station. Such stops are less central and would have a minimal impact on Metro circulation in the event of disruptions. The average degree centrality across the network is 0.0476.

For closeness centrality, the highest value is 0.1173 at Senhora da Hora, a stop strategically located near the center of the Metro network. Its position allows for quick and efficient access to other parts of the system, making it a pivotal node.

On the other hand, the lowest closeness centrality value is 0.0459 at Fânzeres, another end-of-line station. This suggests that passengers at this stop must travel through more connections or longer travel paths to reach other stops compared to more centrally located stops.

Moving on to betweenness centrality, the maximum value is 0.6615, also at Senhora da Hora, further highlighting its importance in the network. This stop serves as a key connector, enabling shorter paths between other stops. In contrast, the minimum betweenness centrality is close to zero, found at end-of-line stops like Fânzeres. These stops do not lie on any shortest path in the network, reducing their role as connectors.

The presence of communities within the Metro network was identified using the greedy modularity algorithm. These communities represent clusters of stops with stronger internal connections and can be visualized in Figure 7.

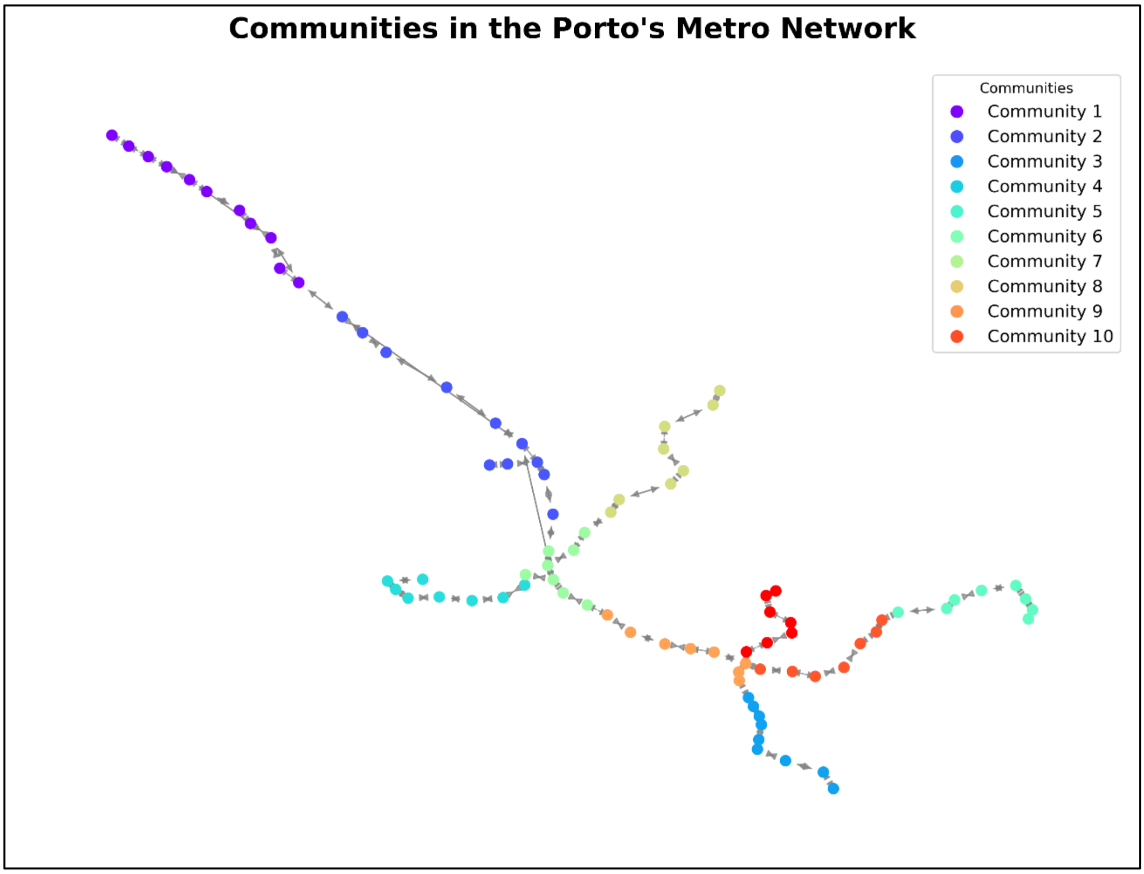

FIGURE 7 – Communities in the Porto's Metro Network

Lastly, when examining whether the Metro network is assortative or disassortative, it was found to have a positive degree of 0.2238, indicating that the network is assortative. This means that stops with similar characteristics, such as a high number of connections, are more likely to be connected to one another. For example, stops with a higher degree of connectivity, such as Lapa, Carolina Michaelis, and Casa da Música, are closely linked to the network's main hub, Trindade. This clustering of well-connected stops highlights the assortative nature of the Metro system.

3.3.2 Overpass Turbo Data

The first step in preparing the data from Overpass Turbo involved uploading the dataset into Jupyter Notebook. Since the data was in GeoJSON format, the geopandas library was imported to facilitate reading and handling the file. This dataset contained 2 531 Points of Interest (POIs), each with a unique and no null geospatial location. Once uploaded, data cleaning was performed to make the dataset more accessible and readable for the analysis.

The initial step in the cleaning process involved examining the unique values of the columns to be used in the analysis (id, addr: city, amenity, name, tourism, and geometry). Upon inspecting the unique values in the addr: city column, it was observed that the dataset included POIs from 40 different cities. However, some of these cities, such as Oliveira de Azeméis, Vale de Cambra, and Caldas de Vizela, were located too far from Porto and the Metro of Porto network. As a result, these cities, along with 25 others, were excluded from the analysis.

After filtering the data, the remaining cities included Porto, Matosinhos, Vila Nova de Gaia, Póvoa de Varzim, Vila Chã, Vila do Conde, Cedofeita, (Santo Ildefonso, Sé, Miragaia, São Nicolau e Vitória), Maia, São Mamede de Infesta, Campanhã, Moreira, and Perafita. This narrowing of the dataset reduced the number of POIs from 2 531 to 177, ensuring that only relevant locations were considered for the analysis. The dataset includes a column named Tourism that indicates the category of each POI, which allows for an analysis of the types of tourism-related POIs present.

These categories include attractions, hotels, museums, hostels, guest houses, apartments, galleries, wine cellars, information centers, and artwork. Among these, most POIs were related to accommodation, with 51 hotels, 34 hostels, and 32 guest houses. Museums accounted for 29 POIs, making them the next most prominent category.

Additionally, a code was executed to determine the Coordinate Reference System (CRS) of the POIs' geospatial data. The CRS was identified as EPSG:4326, which is commonly used for geographic coordinates in latitude and longitude. Finally, to visualize the distribution of POIs, a map was created using the Matplotlib library. On this map the POIs are represented as green circles, providing a clear visual representation of their locations as shown in Figure 8.

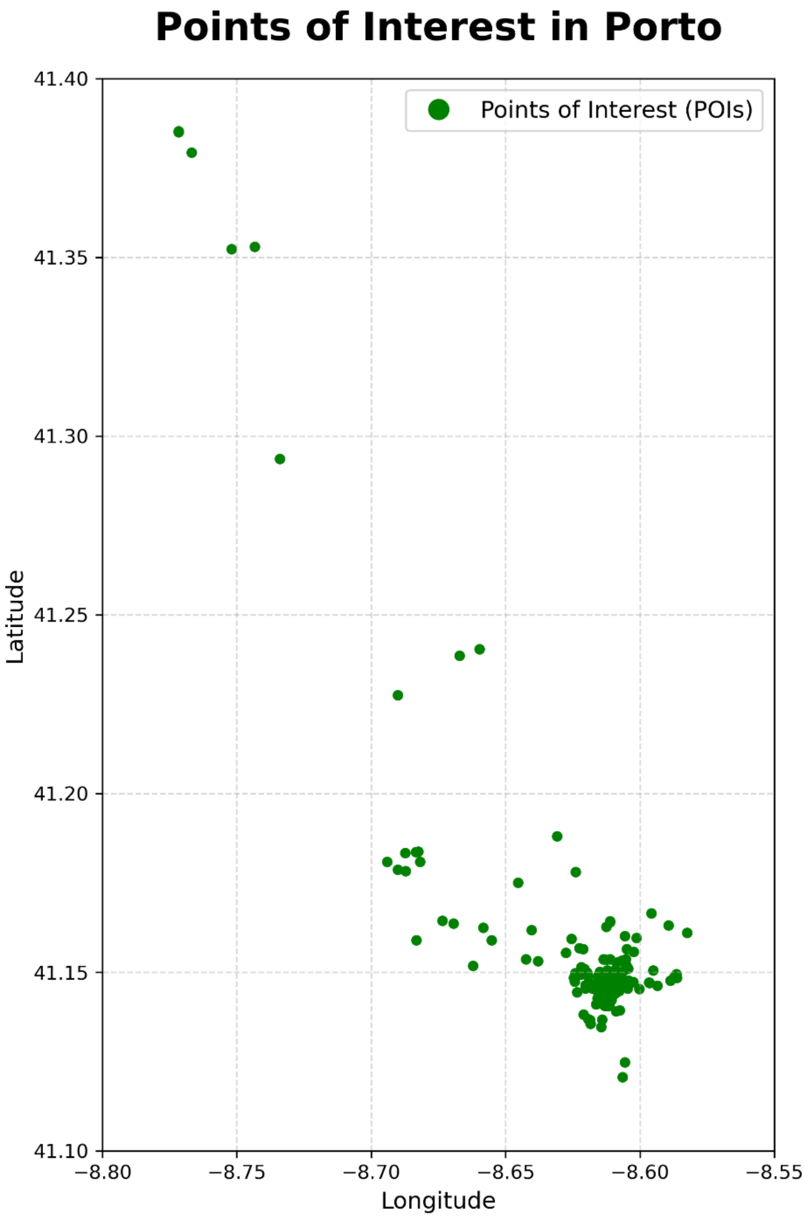

FIGURE 8 – Points of Interest in Porto

3.4 Modeling

In the modeling phase, the objective is to identify and apply the most appropriate techniques to the given data. This ensures the ability to evaluate the performance and outcomes of the applied techniques during the subsequent stages of the project. (Bemthuis et al., 2024)

To address the first question, "How resilient is Porto's Metro system?", two techniques were employed: node removal impact on the LCC size and percolation analysis. These methods were chosen to evaluate the Metro network's resilience when nodes or edges are removed. Combined, they provide a comprehensive view identifying both specific critical stops and the overall resilience of the Metro system. The first technique involves simulating node removal using three distinct strategies: removal based on betweenness centrality, removal based on degree centrality, and random removal. Each strategy helps to reveal how the LCC size responds to the removal of critical nodes versus random failures. On the other hand, percolation analysis evaluates the impact of removing fractions of nodes and edges on the network's overall resilience. Unlike the first technique, which focuses on the impact of removing specific nodes, percolation analysis provides a broader view of how the entire network behaves as groups of nodes or edges are incrementally removed. Nodes are removed using three strategies: random removal, betweenness centrality, and degree centrality, while edges are removed using only random removal and betweenness centrality. Additionally, percolation thresholds are calculated. These thresholds indicate the fraction of nodes or edges that must be removed for the LCC to drop below a critical percentage of the original network size. By examining thresholds at 70%, 50%, and 30%, it is possible to identify how quickly the Metro system becomes fragmented under different failure scenarios. (Lekha & Balakrishnan, 2018) (Rong et al., 2022)

To address the second question, "To what extent are Points of Interest (POIs) in Porto accessible via the Metro?", accessibility was calculated based on the distances between POIs and Metro stops. To perform these calculations, the POIs and the Metro datasets were merged, ensuring that their CRS were compatible and that their geometry types matched. The definition of accessibility in this study was inspired by the 15-minute walking concept discussed in Aparicio et al (2024) and supported by Cronkleton (2019) findings on the average adult walking speed. According to these studies, a 15-minute walk corresponds to a maximum distance of approximately 1 287 meters. However, to provide a more comprehensive analysis, additional thresholds were also considered: 500 meters, 1 000 meters, and 1 500 meters. For each threshold, a POI is considered accessible if it falls within the specified distance being analyzed. This approach acknowledges the variability in walking speed, which can be influenced by factors such as age, physical health, road conditions, and characteristics.

3.5 Evaluation

The main objective of this phase is to evaluate how well the chosen techniques perform in addressing the research questions outlined in the introduction chapter. This involves critically analyzing the technique's results to determine its effectiveness and accuracy in achieving the project's goals. (Bemthuis et al., 2024)

In addition to assessing the performance, this phase also seeks to identify any limitations or challenges encountered during the modeling process. Highlighting these limitations is essential for ensuring a more precise and unbiased analysis, as it provides context for the results and suggests areas for improvement.

The detailed evaluation of the techniques and calculations applied, along with a discussion of the findings, will be presented in the results and discussion chapter, offering a comprehensive overview of the techniques' performance and their alignment with the research questions.

3.6 Deployment

While the deployment phase will not be carried out as part of this project, it is important to outline its role if the model and insights were to be deployed in a real-world scenario. The deployment phase would involve integrating the results and findings into the system of "Metro do Porto" or other companies within the tourism sector to optimize processes like urban planning or transportation management. A plan for monitoring and maintaining the system would ensure its long-term reliability, while a final report would document the methodologies and applications of the companies that acquired the model. Finally, a project review would evaluate the outcomes and identify areas for improvement, ensuring alignment with the project's objectives. (Bemthuis et al., 2024)

4. RESULTS AND DISCUSSION

4.1 How Resilient is Porto Metro System?

As referred in the modeling phase, two techniques were used to assess the resilience of Porto's Metro network. The first technique involved simulating node removal to analyze its impact on the LCC size under three distinct strategies: betweenness centrality, degree centrality, and random removal. The impact of these strategies helps evaluate the network's ability to maintain connectivity when critical or random nodes are removed. The results are the following:

Removal based on betweenness centrality:

This strategy prioritizes removing the nodes with the highest betweenness centrality, as these are crucial for enabling the flow of passengers across the network. The results in Table 1 reveal that even the removal of the single most critical stop causes a significant drop in the LCC size. After removing just 10 of these key stops, the LCC size declines sharply to 20%. This highlights the Metro system's heavy reliance on these connector nodes and underscores its vulnerability to disruptions targeting these critical stops.

TABLE 1 – LCC Size After Betweenness Centrality Removal

Step	Stop	Betweenness Centrality	LCC Size (Normalized)
1	Senhora da Hora	0.6615	0.4824
2	Trindade	0.5823	0.4000
3	Sete Bicas	0.5049	0.4000
4	Viso	0.5034	0.4000
5	Ramalde	0.5014	0.4000
6	Francos	0.4989	0.4000
7	Casa da Música	0.4957	0.4000
8	Carolina Michaelis	0.4920	0.4000
9	Lapa	0.4877	0.4000
10	Pedras Rubras	0.3854	0.2000

Removal based on degree centrality:

This strategy involves removing the nodes with the most direct connections (highest degree centrality). The results in Table 2 show that the LCC size also decreases significantly when critical nodes are removed, but the rate of decline is slower compared to the betweenness centrality strategy. This suggests that the Metro network is slightly more robust to disruptions affecting highly connected stops than to disruptions targeting high-betweenness stops. Even so, the LCC size still drops to lower than 20% after removing the top 10 highly connected stops.

TABLE 2 – LCC Size After Degree Centrality Removal

Step	Stop	Degree Centrality	LCC Size (Normalized)
1	Senhora da Hora	0.0952	0.4824
2	Trindade	0.0952	0.4000
3	Pedras Rubras	0.0952	0.2000
4	Portas Fronhas	0.0952	0.2000
5	Varziela	0.0952	0.2000
6	Vila do Conde	0.0952	0.2000
7	Fonte do Cuco	0.0714	0.1765
8	Mindelo	0.0714	0.1765
9	Modivas	0.0714	0.1765
10	Verdes	0.0714	0.1765

Random removal:

Unlike the targeted strategies, random removal does not focus on specific node importance, such as betweenness or degree centrality. As expected, the impact on the LCC size is much less severe. The results in Table 3 show a gradual decline in connectivity, demonstrating the network's resilience to random disruptions. Even after removing 10 stops, the LCC size remains at 42.35%, which is significantly higher than the 20% and 17.65% observed in the other two strategies. This indicates that the Metro network is more robust to random failures than to targeted attacks on critical stops.

TABLE 3 – LCC Size After Random Removal

Step	Stop	LCC Size (Normalized)
1	Santo Ovídio	0.9529
2	Francos	0.5529
3	Brito Capelo	0.5176
4	Custóias	0.5059
5	Alto da Pega	0.4941
6	Vasco da Gama	0.4235
7	Mercado	0.4235
8	Heróismo	0.4235
9	Salgueiros	0.4235
10	Estádio do Mar	0.4235

To provide a clearer understanding of the results across the 3 strategies, a plot was generated, as shown in Figure 9. This plot underscores the Metro network's significant vulnerability to targeted disruptions, particularly when key nodes are removed based on betweenness or degree centrality, compared to random node removals. The plot clearly illustrates how critical certain stops are to maintaining overall connectivity and highlights the substantial impact that the loss of these strategic nodes can have on network resilience.

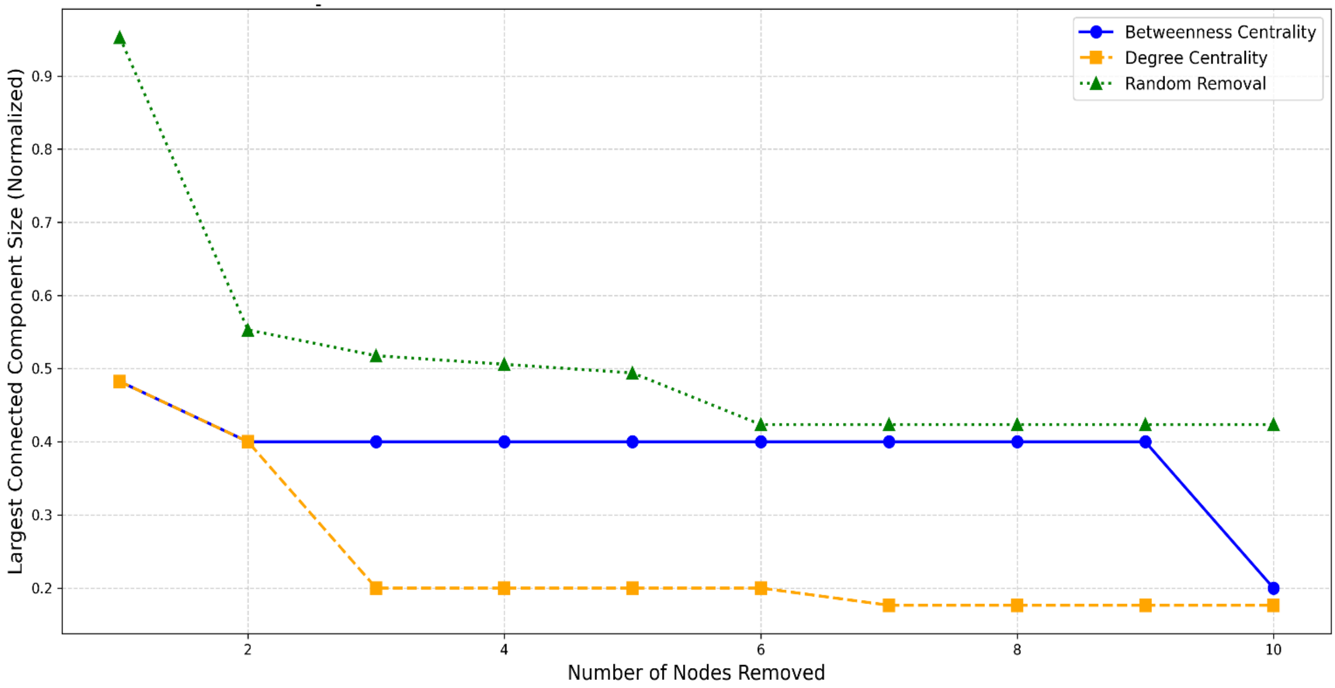

FIGURE 9 – LCC Size After the 3 Strategies

The second technique also focused on node removal, but this time, it was done through percolation analysis. This approach evaluates the resilience of the network by progressively removing fractions of nodes and analyzing the impact on the LCC size. The analysis was performed using three strategies: random removal, degree centrality removal, and betweenness centrality removal. The results are as follows:

TABLE 4 – LCC Size After Percolation Analysis

Fraction removed	LCC Size (random)	LCC Size (degree)	LCC Size (Betweenness)
0.047059	0.800000	0.200000	0.400000
0.094118	0.694118	0.176471	0.400000
0.141176	0.694118	0.129412	0.200000
0.188235	0.494118	0.129412	0.129412
0.235294	0.482353	0.129412	0.117647
0.282353	0.376471	0.129412	0.105882
0.329412	0.329412	0.129412	0.094118
0.376471	0.152941	0.129412	0.094118
0.423529	0.152941	0.129412	0.094118
0.470588	0.129412	0.129412	0.070588
0.517647	0.047059	0.129412	0.070588
0.564706	0.047059	0.129412	0.047059
0.611765	0.035294	0.129412	0.047059
0.658824	0.023529	0.129412	0.047059
0.705882	0.023529	0.129412	0.035294
0.752941	0.023529	0.082353	0.035294
0.800000	0.023529	0.047059	0.035294
0.847059	0.023529	0.047059	0.023529
0.894118	0.023529	0.047059	0.023529
0.941176	0.011765	0.011765	0.011765
0.988235	0.011765	0.011765	0.011765
1.000000	0.000000	0.000000	0.000000

From Table 4, it is clear that similar to the previous analysis, the removal of random fractions of stops in the early stages has a minimal impact on the resilience of the Metro system. The system only becomes unreliable in maintaining passenger flow when approximately 24% of the network is removed. However, the results for degree centrality and betweenness centrality removal reveal a much more critical impact on the LCC size. With the removal of just 4.7% of the Metro system, the LCC drops drastically to 20% for degree centrality removal and 40% for betweenness centrality removal. While both strategies demonstrate significant impacts, the betweenness centrality removal results in a faster and more pronounced decline in the LCC size compared to degree centrality removal. This impact and behavior of the LCC size can also be visualized in Figure 10, where the difference between the three removal strategies is clearly illustrated.

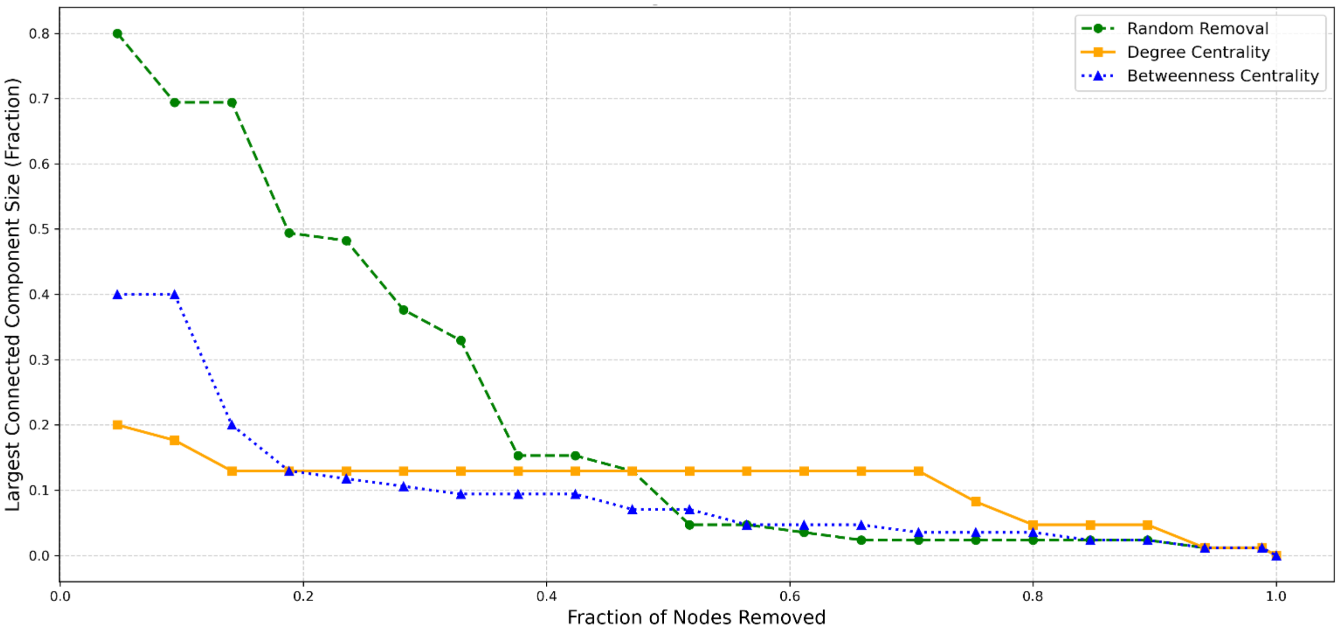

FIGURE 10 – LCC Size After Percolation Analysis

Percolation thresholds were calculated to assess the resilience of Porto's Metro network further. This analysis helps determine the fraction of nodes that need to be removed before the LCC shrinks below critical thresholds of 70%, 50%, and 30% of its original size. The results are summarized in Table 5, showing the percolation thresholds for 3 different removal strategies: random removal, degree centrality removal, and betweenness centrality removal.

TABLE 5 – LCC Size After Percolation Thresholds Analysis

Threshold	Random Removal	Degree Centrality	Betweenness Centrality
70%	0.094118	0.047059	0.047059
50%	0.188235	0.047059	0.047059
30%	0.376471	0.047059	0.141176

The results show that random node removal has a significantly lower impact on network connectivity compared to the targeted removals. The network maintains at least 70% of its connectivity until approximately 9.4% of nodes are randomly removed. However, when nodes are removed based on degree or betweenness centrality, the LCC drops below 70% after removing just 4.7% of nodes, highlighting the critical role of these highly connected or strategically positioned stops. A similar pattern emerges at the 50% and 30% LCC thresholds.

These findings reinforce the idea that Porto's Metro system is more vulnerable to targeted attacks on key nodes than to random failures. While random disruptions gradually weaken the system, strategic removals rapidly fragment the network, threatening overall connectivity.

4.2 To What Extent Are POIs in Porto Accessible Via the Metro?

As referred to in the modeling phase, the accessibility of POIs was determined by calculating the distance between Metro stops and POIs across four thresholds: 500 meters, 1 000 meters, 1 287 meters (the primary threshold based on the 15-minute walking standard), and 1 500 meters. The results of this analysis are summarized in the following Table:

TABLE 6 – Percentage of Accessible POIs

Threshold (meters)	Accessible POIs	Inaccessible POIs	Accessible %
500	127	50	71.75
1 000	156	21	88.13
1 287	163	14	92.09
1 500	166	11	93.78

The results indicate a clear trend: as the threshold distance increases, more POIs become accessible via the Metro system. At the primary threshold of 1 287 meters, only 14 POIs (7.91%) remain inaccessible, meaning the Metro network effectively covers 92.09% of the POIs within this walking distance. However, when the threshold is decreased to 500 meters, the number of inaccessible POIs rises significantly to 50, covering only 71.75% of POIs. This analysis demonstrates the strong coverage provided by Porto's Metro system at larger walking distances but highlights gaps in accessibility at shorter thresholds. The findings suggest that while most POIs are well-connected at a 15-minute walking distance, those POIs that remain inaccessible may require alternative transportation options or improvements to the Metro system to ensure bigger coverage of the city. Additionally, the relatively small increase in accessibility between 1 287 meters and 1 500 meters (from 92.09% to 93.78%) suggests that there is less impact as the threshold increases. This indicates that most of the network's coverage is already optimized within a 15-minute walking distance, and further extensions would only marginally improve accessibility. Visualizations such as the one shown in Figure 11 were generated to provide better insight, highlighting the exact locations of inaccessible POIs.

In this chapter, only the plot for the primary threshold of 1 287 meters is presented, while the visualizations for the other thresholds are included in the annexes in Figure 12.

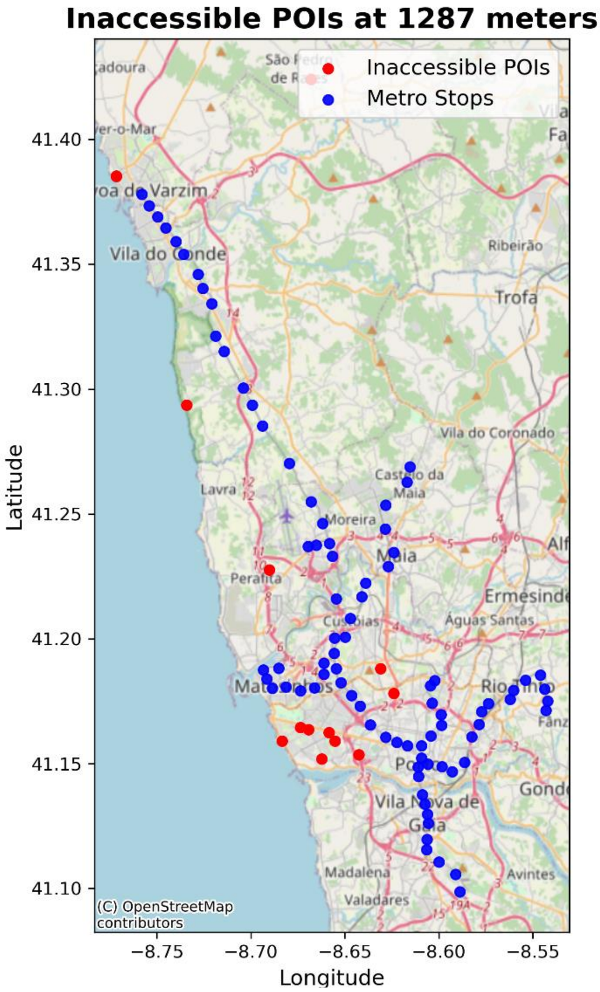

FIGURE 11 – Map of Inaccessible POIs and Metro stops

This visualization not only illustrates the spatial distribution of inaccessible POIs but also helps to identify areas that could benefit from improvement or Metro system expansion. From the map, zones like Foz and Massarelos stand out as key areas with accessibility gaps. By addressing these gaps, the Metro network could improve accessibility, attract more riders, and improve connections to POIs. However, it is important to note that these suggestions do not account for other factors, such as the availability of space for new stops, the type of soil, or the budget constraints for implementing additional Metro stops in Porto.

4.3 Additional Analysis

To complete this analysis, the calculation of the resilience of POIs accessibility was also incorporated. In practical terms, this involved examining how access to POIs is affected when metro stops are removed based on the three removal strategies: degree centrality, betweenness centrality and random removal. For each strategy, ten stops were removed, allowing for a comparative assessment of their impact on accessibility. The results are the following:

TABLE 7 – Percentage of Accessible POIs After Stops Removal

	Accessibility		
Threshold (meters)	**Random Removal**	**Degree Centrality**	**Betweenness Centrality**
500	70.62	70.62	68.36
1 000	87.57	88.14	84.18
1 287	92.09	91.52	89.27
1 500	93.79	93.79	90.96

The results show that removing stops based on betweenness centrality creates the biggest drop in accessibility across all distance thresholds. This suggests that the metro stops with high betweenness centrality play a crucial role in maintaining access to POIs, as they serve as key connections within the network. When these stops are removed, accessibility is disrupted more severely than when stops are removed based on degree centrality or at random. In comparison, the impact of degree centrality and random removals on accessibility remains quite similar, with both strategies showing a less drastic change. Additionally, as expected the results reveal that POIs accessibility is more resilient at higher distance thresholds. This means that when considering a bigger range, the accessibility to POIs is less affected by metro stop removals. In comparison, when considering shorter distances, the effect is much more pronounced, making accessibility more fragile under disruptions. A real-world example of this occurrences could be the comparison the impact of disruptions of the metro stops Senhora da Hora (which has high betweenness centrality), Trindade (which has high degree centrality), and Santo Ovídio (a random removed stop) due to an accident or a natural disaster such as a flood. The effect on accessibility to POIs is not the same since the first two stops are hubs and Santo Ovidio is an end of line stop, which does not affect significantly the rest of the other lines and stops. When comparing the effect of removing Senhora das Hora and Trindade, the first one has a bigger impact since Trindade is already closer to more POIs than Senhora da Hora. These findings suggest that stops with high betweenness centrality should be prioritized in urban planning and transportation management to minimize the risk of loss of accessibility. One possible solution could be implementing a circular metro line that connects more POIs and integrates with other lines, ensuring that if a key stop is affected, passenger flow and accessibility remain safe.

This analysis revealed a similar pattern to the Zhengzhou Metro study, highlighting that more central stops play a critical role in maintaining network connectivity and overall Metro resilience.

5. CONCLUSION

This MFW was produced to answer two research questions: "How resilient is Porto's Metro system?" and "To what extent are Points of Interest (POIs) in Porto accessible via the Metro?". To approach these research questions, network science was applied and combined with various other studies such as network analysis, percolation analysis, and accessibility. This combination provided the study of the Metro's weaknesses and strengths while also examining its accessibility to POIs around the city of Porto.

To address the first research question, the simulation of node removal was considered using techniques based on betweenness centrality, degree centrality, and random removal, followed by a percolation analysis. The results indicate that the Metro system is highly susceptible to disruptions when key stops are deliberately removed compared to random stop removals. In a real-world scenario, this is akin to removing a critical station such as Stop Senhora da Hora, which has the highest betweenness and degree centrality, versus removing a less significant stop like Stop Santo Ovídio. The impact differs significantly, as the first one serves at least five Metro lines, while the second is an end-of-line station with only one line passing through. Consequently, the removal of Senhora da Hora has a much greater effect on passenger flow.

In summary, the Metro system is highly vulnerable to targeted removals of key stops but remains relatively resilient to the removal of less critical, end-of-line stops. To address the second research question, the distance between the Metro stops and the tourism POIs was calculated, followed by the percentage of accessibility within four different distance thresholds: 500, 1 287, 1 000 and 1 500 meters. While the focus was on the 1 287-meter distance thresholds, additional thresholds were included since not every person has the same walking speed. The results indicate that 92.09% of POIs are accessible within a threshold of 1 287 meters, and even within a threshold of 500 meters, 71.75% of POIs remain accessible. This confirms that the Metro system provides strong coverage for key tourism locations. Additionally, the analysis highlights specific areas, such as Foz and Massarelos, where Metro accessibility could be improved. In terms of limitations during the process of this MFW, it can be the fact that both datasets - the Metro GTFS and OverPassTurbo - didn't meet the demand of passengers and tourists in each dataset. This missing information would have been valuable in understanding the actual passenger flow across Metro lines and assessing the real impact on daily circulation in terms of Metro resilience. Additionally, for the POIs analysis, having data on the number of visitors per location would have provided deeper insights into the potential impact on tourism accessibility and demand.

Since no project or study is perfect at first, some future improvements could be implemented to expand the analysis and better the results. One potential development could be the simulation of new Metro stops to establish a circular line covering a significant portion of the city. This could serve as a solution to mitigate the impact of both targeted and random disruptions, ensuring a more resilient Metro system and reducing disruptions to passenger flow.

In conclusion, this study provides valuable insights into Porto's Metro system, highlighting both its strengths and areas for improvement. By strengthening critical stops and improving accessibility, the Metro can become a more resilient, efficient, and inclusive transport system, ensuring its continued role as the backbone of urban mobility in Porto.

REFERENCES

Amghar, R., Jaber, S., Moghaddam, S. M. H. M., Bhouri, N., & Ameli, M. (2023). Resilience as a Service for Transportation Networks: Definition and Basic Concepts. *Transportation Research Record Journal of the Transportation Research Board*. https://doi.org/10.1177/03611981231170180

Aparicio, J. T., Arsenio, E., Santos, F. C., & Henriques, R. (2024). Walkability defined neighborhoods for sustainable cities. *Cities, Volume 149*. https://doi.org/10.1016/j.cities.2024.104944

Bemthuis, R. H., Govers, R. R., & Asadi, A. (2024). A CRISP-DM-based Methodology for Assessing Agent-based Simulation Models using Process Mining (No. arXiv:2404.01114). *arXiv*. https://doi.org/10.48550/arXiv.2404.01114

Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E., & Havlin, S. (2010). Catastrophic cascade of failures in interdependent networks. *Nature*. https://doi.org/10.1038/nature08932

Costa, C. J., & Aparicio, J. T. (2020). POST-DS: A Methodology to Boost Data Science. *2020 15th Iberian Conference on Information Systems and Technologies (CISTI)*. https://doi.org/10.23919/CISTI49556.2020.9140932

Costa, C.J., Aparicio, J.T. (2021). A Methodology to Boost Data Science in the Context of COVID-19. In: Arabnia, H.R., et al. Advances in Parallel & Distributed Processing, and Applications. Transactions on Computational Science and Computational Intelligence. Springer, Cham. https://doi.org/10.1007/978-3-030-69984-0_7

Cronkleton, E. (2019, March 14). Average Walking Speed: Pace, and Comparisons by Age and Sex. *Healthline*. https://www.healthline.com/health/exercise-fitness/average-walking-speed

Delgado Rodriguez, S. (2024). What is GTFS? - General Transit Feed Specification. *GTFS*. https://gtfs.org/getting-started/what-is-GTFS/

Derr, A. (2021, April 16). Network Centrality: Understanding Degree, Closeness & Betweenness Centrality [Visible Network Labs]. *Visible Network Labs*. https://visiblenetworklabs.com/2021/04/16/understanding-network-centrality/

Dinh, T. N., Li, X., & Thai, M. T. (2015). Network Clustering via Maximizing Modularity: Approximation Algorithms and Theoretical Limits. *2015 IEEE International Conference on Data Mining (ICDM), 101-110*. https://doi.org/10.1109/ICDM.2015.139

Feng, Q. K., Zhang, D. M., & Huang, Z. K. (2024). Impact of extreme rainfall on Metro network resilience: Shanghai Metro as an example. *IOP Conference Series: Earth and Environmental Science, 1333*. https://doi.org/10.1088/1755-1315/1333/1/012016

G. Lewis, T. (2011). Network Science: Theory and Applications. *Wiley*.

Guida, C. (2020). (PDF) Urban Accessibility: The paradox, the paradigms and the measures. A scientific review. *TeMa - Journal of Land Use, Mobility and Environment, 13, 149-168*. https://doi.org/10.6092/1970-9870/6743

IBM. (2021, August 17). SPSS Modeler Subscription. *IBM*. https://www.ibm.com/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview

Latitude.to. (2024). GPS coordinates of Porto, Portugal. Latitude: 41.1496 Longitude: -8.6110. *Latitude.to, Maps, Geolocated Articles, Latitude Longitude Coordinate Conversion*. http://latitude.to:8080/map/pt/portugal/cities/porto

Lekha, D. S., & Balakrishnan, K. (2018). Attack Vulnerability of Complex Networks in Center-Based Strategies (No. arXiv:1712.04107). *arXiv*. https://doi.org/10.48550/arXiv.1712.04107

Metro do Porto. (2024a). Metro do Porto: Conhecer. *Metro do Porto* https://www.Metrodoporto.pt/pages/307

Metro do Porto. (2024b). Horários, paragens e rotas da Metro do Porto—GTFS Metro do Porto 06-09-2024—Portal de Dados. *Porto Digital*. https://opendata.porto.digital/dataset/horarios-paragens-e-rotas-em-formato-gtfs/resource/f592c53a-e669-4cac-9e28-ab84b87e7f6b

Moore, C., Yan, X., Zhu, Y., Rouquier, J.-B., & Lane, T. (2011). Active Learning for Node Classification in Assortative and Disassortative Networks. *ArXiv, abs/1109.3240*https://doi.org/10.1145/2020408.2020552

OverPassTurbo. (2024). Overpass turbo. *OverPassTurbo*. https://overpass-turbo.eu/index.html

Qi, Q., Meng, Y., Zhao, X., & Liu, J. (2022). Resilience Assessment of an Urban Metro Complex Network: A Case Study of the Zhengzhou Metro. *Sustainability*, *14*(18), 11555. https://doi.org/10.3390/su141811555

Rafaqat, A. (2023). Understanding Point of Interest: Definition and Key Examples. *Xmap*. https://www.xmap.ai/blog/understanding-point-of-interest-definition-and-key-examples

Rong, Q., Zhang, J., Sun, X., & Wandelt, S. (2022). On the estimation of percolation thresholds for real networks. *Chaos, Solitons & Fractals*, *158,* 111968. https://doi.org/10.1016/j.chaos.2022.111968

Ruhnau, B. (2000). Eigenvector-centrality—A node-centrality? *Social Networks. Volume 22, Issue 4, 2000, 357–365*. https://doi.org/10.1016/S0378-8733(00)00031-9

Tan, P. Y., & Ismail, H. N. (2020). Reviews on interrelationship between transportation and tourism: Perspective on sustainability of urban tourism development. *IOP Conference Series: Earth and Environmental Science*. https://doi.org/10.1088/1755-1315/447/1/012065

Wassmer, J., Merz, B., & Marwan, N. (2024). Resilience of transportation infrastructure networks to road failures. Chaos (Woodbury, N.Y.), 34(1), 013124. https://doi.org/10.1063/5.0165839

APPENDICES

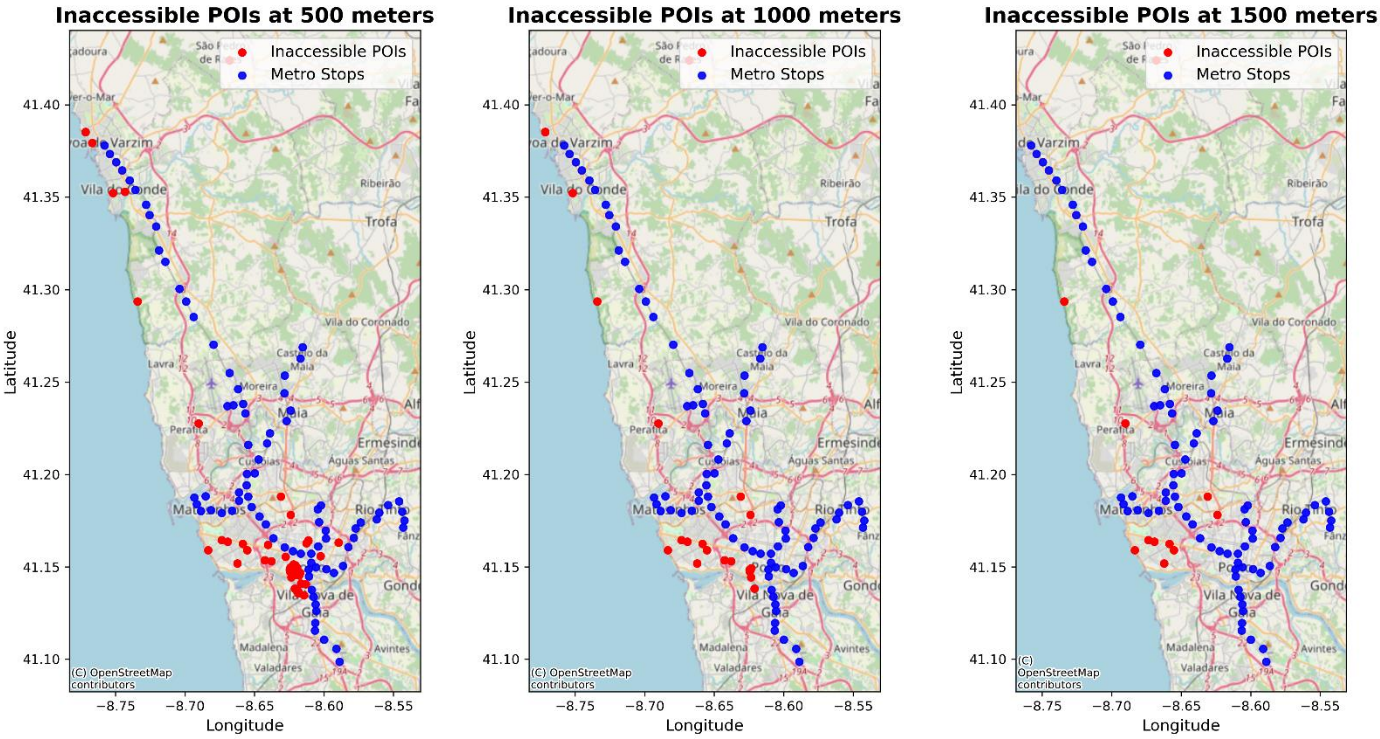

FIGURE 12 – Map of Inaccessible POIs and Metro stops Across Three Distance Thresholds.