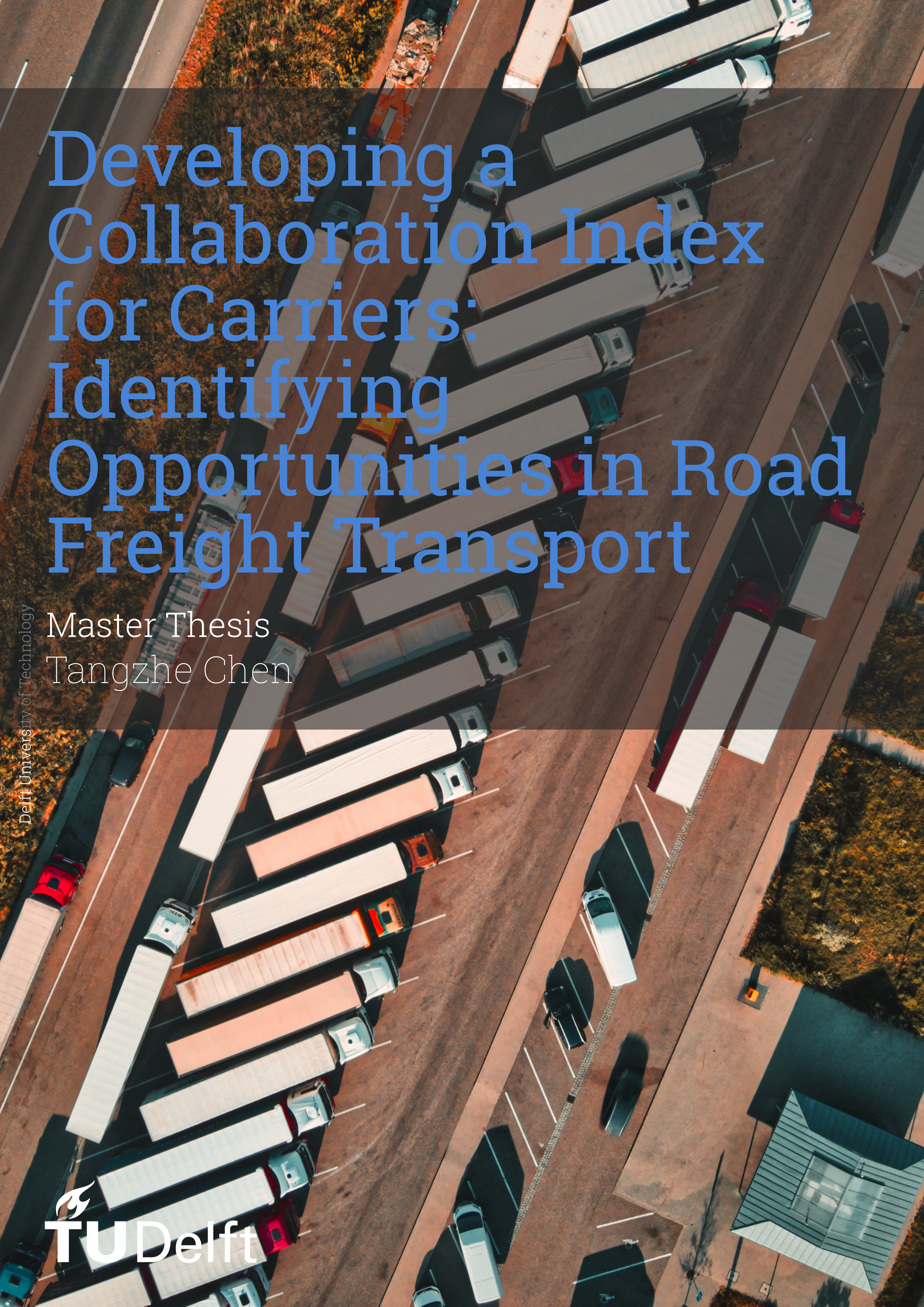
Developing a Collaboration Index for Carriers: Identifying Opportunities in Road Freight Transport

Master Thesis

Tangzhe Chen

Delft University of Technology



Developing a Collaboration Index for Carriers: Identifying Opportunities in Road Freight Transport

by

Tangzhe Chen

to obtain the degree of Master of Science at the Delft University of Technology, to be defended publicly on Monday August 26, 2024 at 2:30 PM.

Student number:	5814162	
Project duration:	February 1, 2024 – August 26, 2024	
Thesis committee:	Prof. dr. ir. Lóránt Tavasszy,	TU Delft
	Dr. ir. Mahnam Saeednia,	TU Delft
	Dr. ir. Ali Nadi Najafabadi,	TU Delft

Cover Image: Download from Pixabay
An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

Summary

In response to the inefficiencies in the European Union's road freight transport, particularly the significant volumes of empty vehicle journeys and their environmental impact, this research proposes an innovative collaboration index for carriers. This index identifies horizontal collaboration opportunities in road freight transport and aims to use this index to help companies find partners to enhance logistics efficiency through horizontal collaboration.

Collaboration index formulation

The framework includes collaboration determination and collaboration index formulation. Among them, the collaboration decision tree determines four modes of collaboration at the trip level from freight transport data, and heuristic local optimization aggregates these into a carrier-level evaluation, formulating a collaboration index between carriers. This collaboration index aims to quantitatively pinpoint the potential for collaboration among companies engaged in road freight transport, addressing a critical gap in the partner selection process for horizontal collaboration, which has been relatively understudied in terms of evaluations based on operational areas such as freight transportation data. Through the formulation of the collaboration index, this study contributes to companies on the business level, giving hints on finding partners while applying horizontal collaboration to reduce their costs and enhance transport efficiency. The collaboration index also helps the coordinator who works as a neutral third party in the collaborative network, using the collaboration index to provide suggestions for their system users.

Application of the collaboration Index

The collaboration index can serve as a key component of the Collaborative Decision Support System utilized by transportation companies to facilitate partnerships. In this system, companies share their data with third-party coordinators, who use this information to assess the collaboration index. This index offers valuable insights into the potential benefits that can be shared between partners in horizontal collaboration. Additionally, the system provides a detailed collaboration plan alongside the index. As illustrated in Figure 1, participating companies can access and review the collaboration index values, enabling them to make informed decisions when evaluating partnership opportunities.

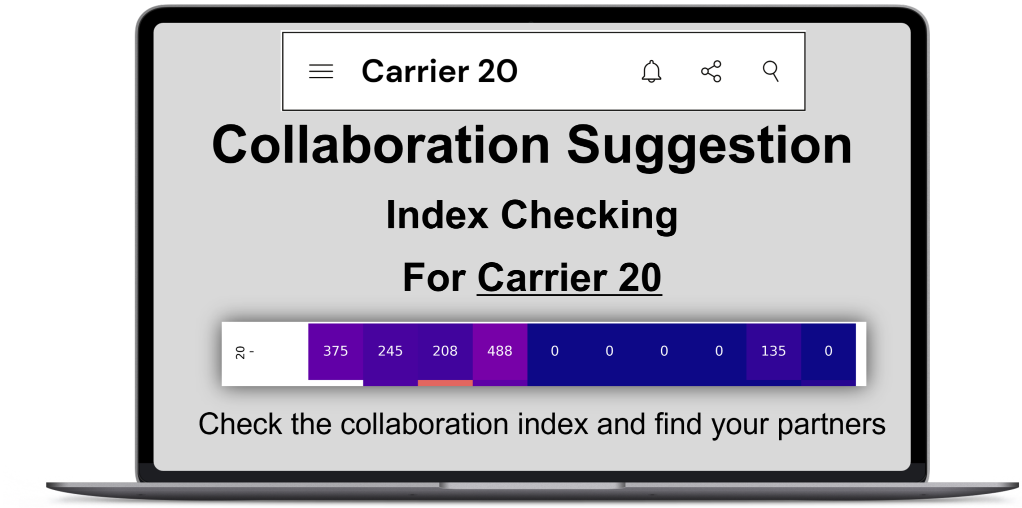


Figure 1: Example: application of collaboration index in the system (UI for carrier 20)

Advantages of the collaboration Index

In this research dataset, companies who participate in horizontal collaboration experienced an average cost-saving rate of 14.4%. The top 25% of these companies achieved a rate of 22.12%. While the geographical similarity of nodes shows a moderately positive correlation with the collaboration index, the vehicle types between partners do not exhibit a linear relationship with the index.

The significance of this research lies in its unique collaboration index formulated from operational data, which not only facilitates the identification of collaboration potentials but also provides insights for the collaborative partner selection process. This index enables carriers to choose collaboration partners by highlighting the optimal benefits of collaboration plans based on all companies' operational data, thereby simplifying the selection process and enhancing the efficiency and effectiveness of horizontal collaboration strategies.

Acknowledgements

I would like to express my deepest gratitude to several individuals who have been instrumental in the completion of this master thesis.

First and foremost, I am very grateful to my thesis chairman, Prof. Lóri Tavasszy, for his insightful feedback, and constructive criticism throughout my research journey. His expertise and encouragement have been important in shaping this work.

I also want to thank to my supervisors, Mahnam Saeednia and Ali Nadi Najafabadi, for their continuous support, and valuable guidance. Their dedication and passion for research have been a constant source of inspiration.

I am deeply thankful to my parents and sister for their unconditional love and support. Their belief in my abilities and their constant encouragement have been the foundation of my academic journey.

I would also like to express my gratitude to my fellow classmates and friends (Yiheng Fang, Ke Zhang, Yunjie Zhang, Ximing Du, Xueqian Gu, Jingmin Liang, Jie Zhan) who have encouraged and supported me throughout this process. Their understanding, encouragement, and moral support have been vital during the challenging moments of this journey.

Lastly, I would like to express my gratitude to everyone who has contributed to my personal and academic growth, making this journey both fulfilling and rewarding. Thank you all for being an integral part of this journey.

Tangzhe Chen
Delft, August 2024

Contents

List of Figures

List of Tables

1 Introduction

1.1 Research context

In 2022, the European Union experienced a significant amount in road freight transport, with a total volume exceeding 13.6 billion tonnes and covering an extensive 1,920 billion tonne-kilometres. Road freight transport constantly increases nowadays. According to the European Union Statistics, the growth of total road freight transport developed at 2.1% every year from 2018 to 2022. A concerning statistic emerged, revealing that at the EU level, 20.2% of the total distances covered were by empty vehicles, posing a considerable challenge to the overall effectiveness. This issue was prevalent across most countries, with high percentages of empty vehicles in both national and international transport. In 2022, at the EU level, 20.2% of the total distances performed by empty vehicles. The Netherlands, in particular, stood out with a staggering 32.6% of empty vehicles in national transport and 22.4% in international transport, surpassing 20% which is mirrored by only a few countries (Eurostat, 2023).

One negative effect of certain inefficient logistics is the cost increase in operation, particularly highlighted by the existence of empty vehicles. This situation occurs when trucks and delivery vehicles travel without cargo, leading to higher operational costs as companies must cover the expenses for these non-revenue-generating trips (McKinnon & Ge, 2006). Moreover, these inefficiencies intensify traffic congestion, which can lead to higher infrastructure maintenance costs.

The shortage of freight transport efficiency not only increases the cost of travelling but leads to other negative consequences as well, for example, the negative impact on the environment of greenhouse gas (GHG) emissions. In 2021, road vehicles contributed about three-quarters of the European Union's transportation emissions. Petroleum-based fuel combustion in road transportation emerged as the largest source of CO_2 emissions in the EU. Passenger cars were responsible for 60% of road transportation CO_2 emissions, while heavy-duty vehicles (trucks and buses) ranked second, emitting over 200 million tonnes of carbon dioxide equivalent annually (200 Mt CO_2e), comprising about one-quarter of road transportation emissions (Statista, 2024). Although heavy-duty vehicles have fewer vehicles on the road, high levels of pollution are emitted by them, emphasizing the need to address their environmental problem. The reduced number of freight transport trips, especially the empty vehicles would help address this problem.

Collaborative logistics has been believed to be one innovative method for tackling the inefficiency in logistics (Amer & Eltawil, 2015; Cruijssen et al., 2007; Y. Wang et al., 2021). Collaborative logistics is a strategy in which companies join together to enhance the effectiveness of their supply chains. This approach contrasts with working isolate accompanied by inefficiency inherent (Ferrell et al., 2020). Horizontal and vertical collaboration are divided by different physical network structures. Horizontal collaboration involves the cooperation of multiple organizations that are either unrelated or competitors. They are at the same logistics level, cooperating for sharing individual information or assets, such as vehicle capacity. On the other hand, vertical collaboration is the joint effort of multiple organizations functioning at various levels of the supply chain. This could include for example manufacturers, distributors, shippers, and carriers, who collaborate by sharing their duties, resources, and data to cater to a similar group of end customers within a specific supply chain (Ferrell et al., 2020; Vargas et al., 2018).

Some research focuses on reviewing the existing literature and published case studies. It has been noted that studies focusing on collaborations often lack grounding in real-world data. Gansterer and Hartl (2020) suggested that the data analysis method such as neural neural network could be used for a deeper understanding of effective strategy applied in the collaboration environment. Key difficulties and obstacles for collaboration are highlighted theoretically including the distribution of revenue, competition law and policy, process coordination, organization compatibility, structure and strategies cooperation (Audy et al., 2012; Vargas et al., 2018). To better understand the nature of barriers, Karam et al. (2021) categorized 31 recognized obstacles into five groups, business models, information exchange, human factors, Collaborative Decision Support Systems (CDSSs), and market influences. The model helped to guide the risks and benefits of implementing collaborations.

Researchers categorized studies related to horizontal collaboration by the context of the problem. The classification includes partner selection and relationship management (Amer & Eltawil, 2015; Essam & Eltawil, 2014). The partner selection research tries to identify and test the strategic fit of companies before they collaborate. Additionally, within the scope of relationship management, issues are further divided into areas such as forecasting, sharing of profits/costs, and collaborative approaches for sharing capacity/demand and routing problems (Ergun et al., 2007; Lai et al., 2022; J. Wang et al., 2018; Y. Wang et al., 2021). The review of literature on the topics of partner selection indicates frequent use of the Analytic Hierarchy Process (AHP) method (Bahinipati et al., 2009; Feng et al., 2010; Naesens et al., 2009). This preference is attributed to the proficiency of AHP in organizing multi-attribute issues into a hierarchical structure, which allows for a detailed examination of each hierarchical level independently. The research on relationship management is much more than the partner selection (Essam & Eltawil, 2014).

The current research primarily focuses on broader perspectives, such as financial aspects, historical sales data, and employee counts. However, there is a noticeable gap in the literature regarding the selection of partners in the operational domain, specifically on trips and relevant freight information. Additionally, while there is a substantial amount of research dedicated to investigating decision-making processes, the step for addressing the identification of potential collaborations among companies through data analysis methods is relatively neglected.

In summary, this research aims to develop an identification collaboration index for road freight transport. The exploration of the collaboration index provides insights between carriers for collaboration.

1.2 Research objective

The objective of this research is to define a collaboration index to identify opportunities for collaboration among carriers. The collaboration index will be developed to assess the potential for collaboration among the companies represented in the dataset.

By developing the collaboration index, the research contributes to guiding companies and authorities for partner selection and evaluation when considering collaborative freight transportation. The quantitative insights induced by the index introduced a direct method considering various attributes and factors for collaboration. Additionally, as a foundation for further collaborative decision-making problem optimization, this research provides a quantitative way for data filtering and data prepossessing.

1.3 Research scope

Building on the concepts outlined in the preceding section, this research aims to address the gap in understanding horizontal collaboration within freight transportation, with a specific emphasis on partner selection. Consequently, the scope of this study is confined to the aspects of collaboration partner selection, and it will not delve into the optimization of various collaboration problems.

This research investigates under the assumption that all companies in the dataset are equal in terms of background, omitting factors such as financial status, number of employees, and other criteria typically considered in current partner selection studies. The focus of this study is exclusively on freight transportation data.

Also, the collaboration index developed in this research is applied at the company (carrier) level, signifying that it represents an aggregate measure based on the cumulative data of trips and tours by each carrier. This index serves as a tool to evaluate the extent of the collaborative potential of a carrier, taking into account the comprehensive logistic activities encompassed in their freight operations. By aggregating individual trips and tours, the index offers an overall insight into the collaboration potential of multiple companies in the dataset.

1.4 Main research question & sub-questions

Based on the objective introduced in section 1.2, the main research question is defined as follows which addressed the objectives:

How can the 'collaboration index' be defined to assess the potential for collaboration among companies involved in road freight transport?

The main question is answered by the following sub-questions:

1. How can the 'collaboration index' be used by companies and authorities for partner selection and evaluation in freight collaboration transportation?
2. What attributes and factors should be included in the 'collaboration index' to reflect the potential for collaborative logistics accurately?
3. How can the 'collaboration index' be formulated to quantitatively measure the potential for collaboration?

2 Literature review

This chapter will comprehensively review and introduce relevant literature related to this research. By exploring scholarly works, theories, and studies, this chapter aims to establish a solid foundation, enriching the understanding of the topic and objective for answering the research question.

2.1 Collaborative transport networks

The Collaborative Transport Networks (CTNs) represent an innovative approach to logistics and supply chain management, where a group of competing companies, including shippers and carriers, opt to collaborate in their freight transport operations (Pan et al., 2019). This collaboration aims to design and implement a shared logistics and transport network, optimizing efficiency and resource utilization across the supply chain. By pooling resources and coordinating efforts, CTNs can achieve significant benefits, including lower transportation expenses, better service quality, and improved sustainability. The approach of CTNs fosters a collaborative relationship among partners, transforming competitive dynamics into cooperative strategies to tackle common logistical challenges effectively (Karam et al., 2021).

In Collaborative Transport Networks (CTNs), the companies (partners) have to collaborate by sharing information about their transport orders and available trucks with a central coordinator. This coordinator uses the shared data to identify opportunities for collaboration. The coordinator, acting as a neutral third party, is trusted by the collaborating partners to manage the collaboration impartially and ensure the secure handling of shared information (Audy et al., 2007).

To handle the large number of information shared from the partners, the coordinator need a system for processing. Collaborative Decision Support Systems (CDSSs) are implemented to manage and facilitate these collaborative processes (Gansterer & Hartl, 2018).

2.1.1 Collaborative Decision Support Systems

Collaborative Transport Networks (CTNs) rely on Collaborative Decision Support Systems (CDSSs) for information sharing and the development of collaborative plans. The CDSS typically comprises three main components: a database, computational algorithms, and an interactive dialogue. The overview of the CDSS is shown in Figure 2.1.

The database module collects and retains all information related to logistics, which can be input manually or automatically from partners' transport planning systems. This data is then utilized by the algorithmic module, which analyzes the shared data and makes collaborative decisions, including proposals for exchanges and the distribution of profits among partners (Karam et al., 2021). Compared to non-collaborative efforts, the financial advantage gained from collaboration is termed the benefits of collaboration or profits. CDSSs are frequently employed to allocate these profits among the participating companies using various profit-sharing mechanisms (Guajardo & Rönnqvist, 2016).

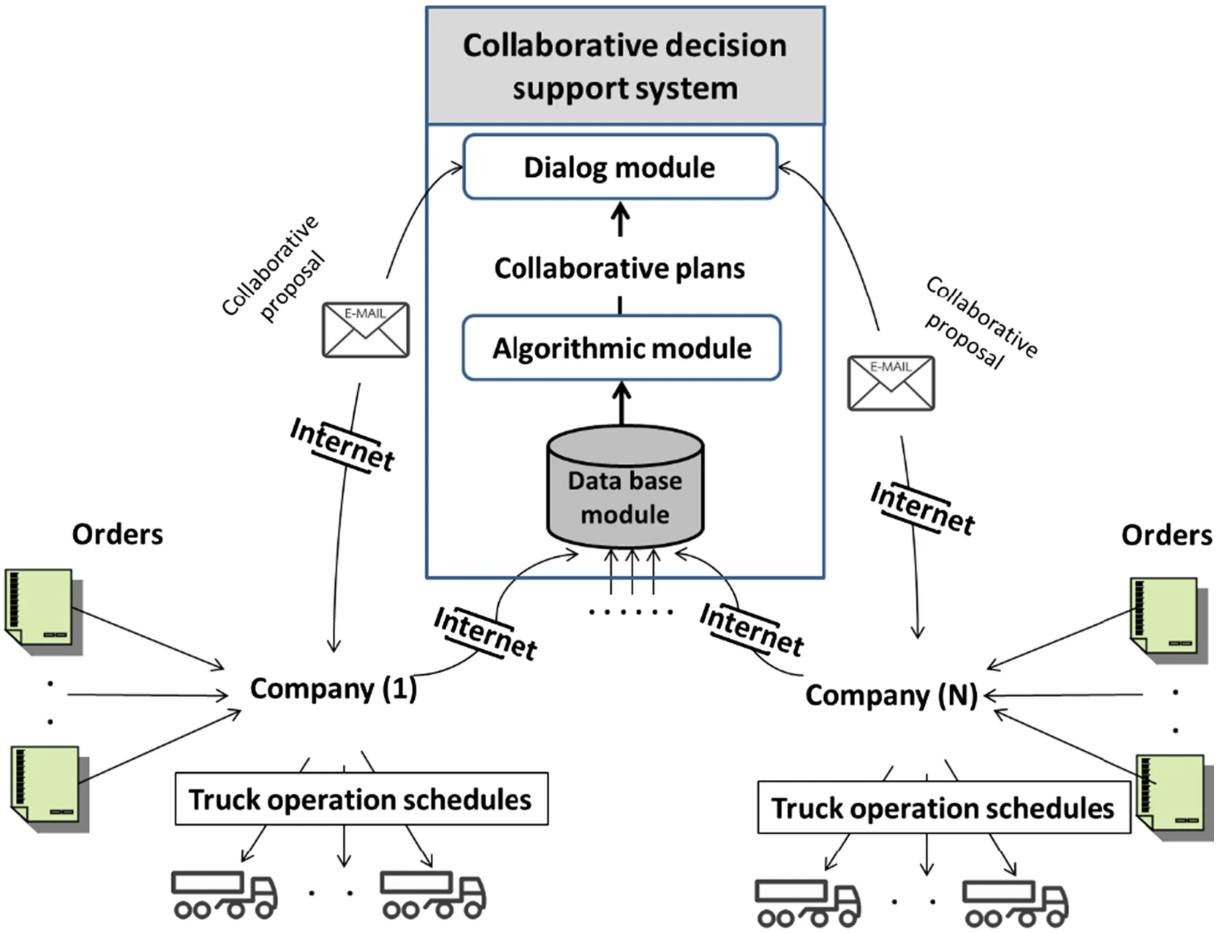


Figure 2.1: Overview of Collaborative Decision Support System (Karam et al., 2021)

There are two main types of algorithms, which are order sharing and capacity sharing. Order sharing requires the order detail information sharing from all the participants including the customer orders with product flows. A central coordinator would optimize orders, and assign the orders matching with participants (Dahl & Derigs, 2011). For the capacity sharing, it requires the sharing of available capacities within the trucks, excluding sensitive customer order information. Capacity sharing is also more applied in companies when considering collaboration (Agarwal & Ergun, 2010).

When considering the partners involved in CDSSs, the user might be carriers, shippers or receivers. The single carrier collaboration represents a peer-to-peer bilateral partnership between carriers, driven by objectives such as reducing transportation costs and enhancing service levels (Pan et al., 2019).

Some research addressed the collaboration partner selection by developing new approaches. Feng et al. (2010) outlined both individual and collaborative attributes and their respective measurements. They proposed a fuzzy multiple attribute decision-making (FMADM) approach to combine assessment data from both sets of utilities, facilitating the comprehensive ranking of potential partners. Awasthi et al. (2016) introduced a pioneer study with fuzzy BOCR-GRA methodology. Through a multi-step process involving criteria identification via BOCR, linguistic assessment with fuzzy logic, and ranking with GRA, this work innovatively navigated the complexities of partner selection under data constraints. The application of a veto mechanism and sensitivity analysis ensured robust and unbiased decision-making, marking a significant advancement in urban logistics planning.

Hernández et al. (2012) addresses a Centralized Carrier Collaboration Multihub Location Problem (CCCMLP) for small to medium-sized less-than-truckload industries. In this problem, a third-party logistics firm aims to establish collaborative consolidation transhipment hubs to create a hybrid hub-and-spoke system that minimizes total collaborative costs for participating carriers. This study considers these factors, unlike previous research that overlooked the location of transfer hubs and connecting routes. Carriers can choose to either collaborate or ship independently, depending on the expected profit margin from collaboration.

Guajardo and Rönnqvist (2016) investigated methods for sharing the benefits of collaboration, identifying various industrial cases and the resulting cost savings. Their findings indicate that collaboration can yield cost savings ranging from 5.99% to 46%. Similarly, Hernández and Peeta (2014) introduced the concept of a degree of collaboration in their model for centralized multiple-carrier collaboration. A higher degree signifies a greater extent of collaboration among carriers, leading to cost savings between 2% and 55%. Recent research by Beliën et al. (2017) demonstrated that horizontal collaboration can reduce transportation costs by 10-15%.

Given the advancements in research methodologies for selecting collaboration partners, as highlighted by Feng et al. (2010) and Awasthi et al. (2016), methodologies need proper factors related being included. The formulation of a collaboration index necessitates the consideration of several factors related, so the review of relevant literature is important. While there are no directly comparable studies currently available for partner selection, exploring optimization papers and topics that are closely related is crucial for grounding our research in a solid theoretical framework.

Main elements were identified and introduced through the classification of potential benefits and barriers (Karam et al., 2021). Information sharing, business models, Collaborative Decision Support Systems (CDSS), market dynamics, and human factors constitute the five primary elements of collaboration as introduced. Among the potential benefits derived from these elements are improved service levels, reduced empty runs, decreased traffic congestion, shorter delivery times, higher resource utilization, reduced lead times, and lower transportation costs. These benefits contribute guide considering various influencing factors to build effective collaborations.

In assessing the optimization model for collaborative logistics, Ghomi et al. (2023) incorporated truck types as a factor. The model accounted for the volume and weight of goods transported by various trucks, as well as the unused weight and volume capacities within these trucks. This analysis differentiated Truck-Load (TL) and Less-than-Truck-Load (LTL) trucks. To evaluate the effectiveness of different collaborative setups, the model considered the average numbers of these truck types and their respective transport distances.

In exploring how logistics platforms can enhance the utilization of vehicle capacity, a study into these platforms reveals key elements in collaboration. Deng et al. (2023) considered the creation of a three-player evolutionary game model (TEGM), which was instrumental in examining the dynamics among participants under various situations. This model's parameters were categorized in alignment with the stakeholders involved. Specifically, for freight carriers, the model takes into account aspects like the likelihood of adopting a shared capacity strategy, associated sharing expenses, payments, potential economic losses, and the resulting income for carriers when different strategies are employed.

When considering related topics, research on ride-sharing bears relevance to this study. Both the ride-sharing and horizontal collaboration thrives on optimizing for efficiency and cost-effectiveness. In ride-sharing, algorithms dynamically match passengers with the most suitable routes and times, while in horizontal freight transport, similar technologies are used to consolidate shipments that share common routes or destinations, maximizing load efficiency and reducing transportation costs. This intersection of insights from the ride-sharing domain also enriches the evaluation framework employed in this research.

Du et al. (2022) embarked on a study to assess the current problems in shared bus mobility through a detailed examination of passenger flows and the recommendation of key indicators. Leveraging data from the Xiong'an New Area between September 2020 and September 2021, their analysis led to the development of a time-series algorithm aimed at identifying ride-sharing trends among demand-responsive (DR) buses. This algorithm facilitated the proposal of indicators for operational efficiency, taking into account factors such as capacity, turnover, and time. The research further delves into the definition of these indicators and their calculation in a specified case study.

Sharing taxis is also a relevant ride-sharing study. To explore the potential demand for taxi ride-sharing and its determinants, Dong et al. (2023) identified and analyzed potentially shareable taxi orders using GPS tracking data from taxis in Chengdu, Sichuan province. They developed a computational method to segregate these shareable orders. Subsequently, they selected 37 potential influencing factors and constructed a multi-factor variable model to quantify their impact on ride-sharing demand. This approach facilitated a nuanced understanding of the factors that drive taxi ride-sharing preferences.

Considering the environmental impact of freight transport, the emission needs to be quantified and used mathematically to show the impact of the influence. Dennis-Bauer and Jaller (2023) built the multiple linear regression model to estimate the public health cost considering various emission factors. J. Wang et al. (2018) used the carbon emission and carbon consumption parameters to express the carbon emission related to energy consumption.

2.2 Conclusion

Following a comprehensive review of existing scientific literature, several notable gaps in knowledge have been identified. Firstly, while Collaborative Transportation Networks (CTNs) offer an innovative approach to managing collaboration, current research predominantly focuses on relationship management within the decision-making process, with less emphasis on partner selection. Secondly, within partner selection, existing studies focused on qualitative aspects such as financial, historical sales data, employee counts, and goal motivations. However, there has been less emphasis on the operational domain, particularly analyses based on freight logistics data. This gap highlights the need for utilizing operational data to identify and optimize collaborative opportunities in freight transportation. Thirdly, the concept of a collaboration index represents an innovative method for quantitatively determining the potential for collaboration between companies, which can be included in the Collaborative Decision Support Systems (CDSS) as an evaluation value to highlight collaboration opportunities. Stakeholders such as coordinators, who act as neutral third parties, maintain the information and offer opportunities to participants. Customers in the CTN engage in the identification process for collaboration, seeking suitable partners. Additionally, the government or authorities may utilize the collaboration index as an indicator to assess the level of collaboration among candidates.

These identified gaps serve as the fundamental for the research objectives of this paper. Table 2.1 provides a comprehensive overview of the distinctions between this research and the relevant literature.

Accordingly, this study aims to delve into the collaboration using freight transport data, which considers several attributes associated with tours and trips. Through the application of the introduced approach, this research seeks to develop an effective 'collaboration index'. This index will be used as the indicator to assess the potential for collaboration. Subsequently, the collaboration index will be utilized to identify suitable partners and evaluate their compatibility for freight transportation, facilitating contribution to the field of logistics and supply chain management.

Table 2.1: Comparison analysis between the current research and literature

Literature	Main Topic	Type of Collaboration	Methodology
Amer and Eltawil (2015)	Green Collaborative	Partner selection Relationship management	Literature research
Dahl and Derigs (2011)	Cost/profit allocation	Relationship management	Simulation
Y. Wang et al. (2021)	Vehicle routing problem	Relationship management	2E-CMDPVRP
Awasthi et al. (2016)	City logistics planning	Partner selection	Fuzzy BOCR–GRA
Feng et al. (2010)	Codevelopment alliance	Partner selection	FMADM
Bahinipati et al. (2009)	Collaboration evaluation	Partner selection	AHP-FLM
Luan et al. (2022)	Opportunity identification	Partner selection	Trailer capacity graph
Hernández and Peeta (2014)	Multihub Location Problem	Optimization	CCCMLP
Ghomi et al. (2023)	Cross-docking	Optimization	VRP
Du et al. (2022)	Intelligent Vehicle Sharing	Bus sharing	Quantitative design
Dong et al. (2023)	Taxi-ride sharing	Taxi-ride sharing	Quantitative design
This research	**Collaboration index identification**	**Partner selection**	**Quantitative design**

3 Research Methodology

In this chapter, the research methodology is divided into two main sections: the determination of collaboration and the formulation of the collaboration index. The first section introduces a comprehensive decision tree designed to identify and classify four distinct modes of horizontal collaborations at the trip level. This detailed flowchart serves as a systematic tool to distinguish various collaboration scenarios, ensuring a clear categorization.

The second section builds upon the identified collaboration modes, delving deeper into the evaluation and quantification processes. This involves transforming the insight of trip-level collaboration into carrier-level collaboration information. By analyzing the trip-level data, the comprehensive collaboration index that reflects the extent and potential of collaborations among carriers is introduced. This index facilitates a better understanding of efficient collaboration and provides a metric for showing the collaborative potential between partnerships in freight road transportation carriers.

3.1 Collaboration determination

Before starting the investigation into collaboration, it is important to note that extensive literature has provided various complex methods for exploring horizontal collaboration optimization among candidates. However, there is a notable gap in research focusing on the partner selection step, specifically in suggesting which pairs of partners should collaborate. Additionally, due to the unavailability of complete customer data for each carrier, particularly detailed production-consumption data for cargo, this section develops a straightforward decision tree to identify four different modes of collaboration using available freight transportation data.

3.1.1 Assumptions

Before introducing the decision tree, it is essential to outline the specific assumptions made in this study. Firstly, the detailed cargo flow from origin to destination is not considered, allowing for a focus on more generalized aspects of the weight movement. Additionally, this research does not take into account the volume used by the cargo of the vehicles; instead, it primarily considers the weight of the cargo. It is assumed that for independent freight transportation (no collaboration), all cargo carried in a single vehicle during its tour is homogeneous, except for variations in weight that occur during loading and unloading at different nodes. These weight variations, observed before and after operations at each node, indicate whether the vehicle is engaged in either a loading or unloading service at that particular node. This assumption simplifies the complexity of cargo management and focuses the analysis on the critical aspect of weight management.

3.1.2 Example of collaboration

In this section, two examples of collaboration, each incorporating specific assumptions, are introduced to facilitate a deeper understanding and provide insights into how collaboration occurs between two trips managed by different carriers. By exploring these instances, the collaboration in freight transportation becomes clearer, offering a perspective on how such trip-level collaborations are formed and operated.

Example 1

The first example is built under additional assumptions, considering collaboration only between the same origin-destination (OD) trip of the two carriers, without taking into account the depots of either carrier. The compatibility of different cargoes is also not included. This example only focuses on the collaboration solutions, therefore the capacity of both vehicles and the weight of the cargo are not considered.

The importance of identical OD lies in its facilitation of shared capacity between two carriers without the need for additional routing. As depicted in Figure 3.1, carrier A and carrier B have their route along the nodes before collaborating. Both carriers have the same origin-destination (OD) trip, visiting the same origin and destination along their respective routes.

Each carrier has its own transport task, which involves transporting cargo between origin-destination (OD) trips. The transport task can be shared between two carriers only if the cargo is loaded at the ”Origin” node, which allows the other carrier to also carry the cargo on its vehicle. Therefore, the operation at the ”Origin” node determines whether the transport task can be shared. Loading at the ”Origin” node indicates the potential for sharing the transport task, whereas unloading does not.

Carriers can share capacity on the same OD trip by having one carrier handle the other’s transport task. Consequently, the other carrier does not need to arrive at the ”Origin” and can instead proceed directly to the ”Destination” from the node preceding the ”Origin”. It is also called as ”New route”. This arrangement saves the OD trip for the carrier.

So for carrier A and carrier B, when they are available for collaboration, two possible modes for collaboration are ”carrier A carry cargo for B”, and ”carrier B carry cargo for A”. The routes after collaboration for both two solutions are also shown in Figure 3.1.

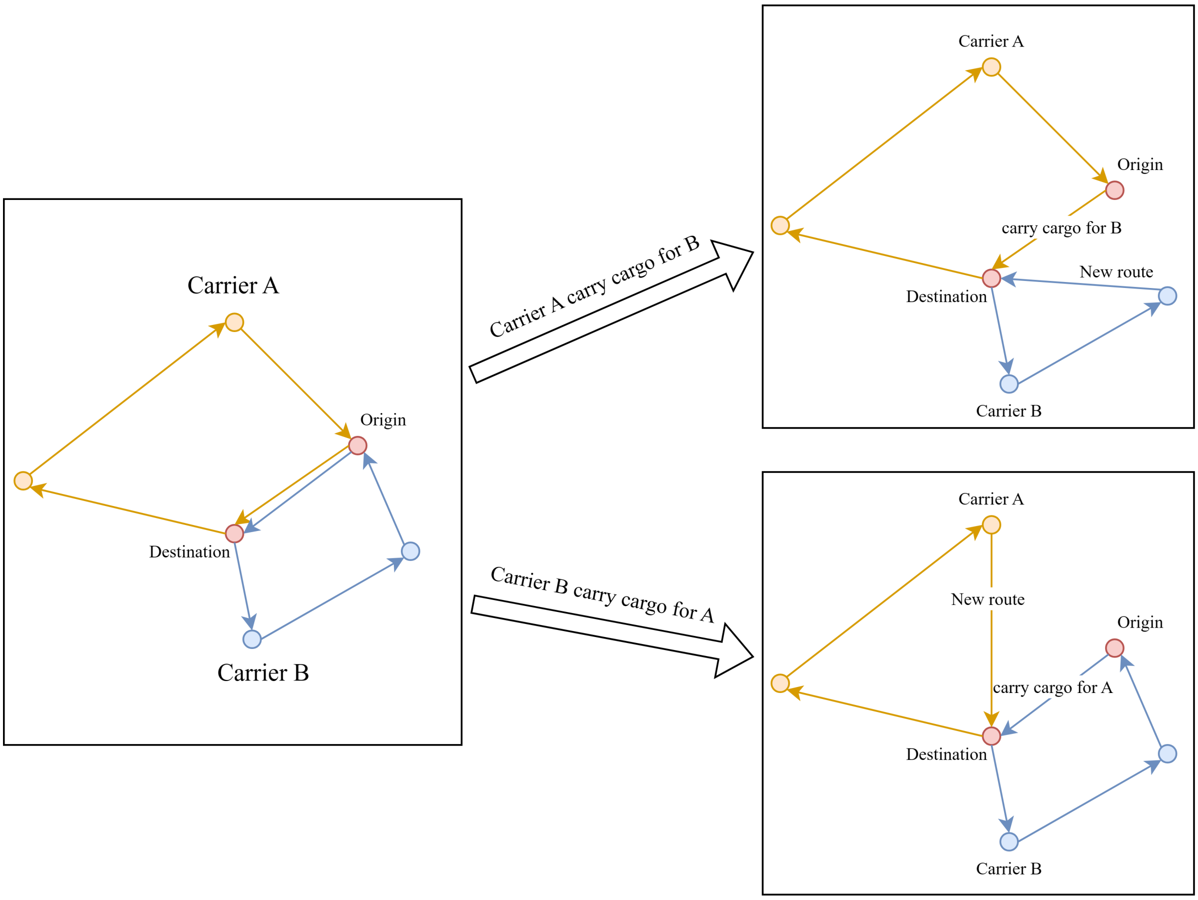


Figure 3.1: The collaboration solutions for the same OD trips from carrier A and carrier B (example 1)

Example 2

The second example is built under the assumption that the depots of neither carrier are considered, the compatibility between cargoes is not taken into account, and the capacity of the vehicles is not included. However, in this example, one assumption changes: the two carriers do not have the same origin-destination (OD) trip. Instead, a certain range between their OD trips is considered, allowing for the reallocation of tasks between them.

In this general example where reallocation is considered, a specific range is established for both origin (ORIG) and destination (DEST) pairs. As illustrated in Figure 3.2, carriers A and B have trip pairs where both ORIG and DEST are located within this predefined range. In this example, the origins and destinations for both carrier A and carrier B are the points being investigated, where the origin pairs and destination pairs fall within a certain range.

If carrier A is responsible for covering the transport task typically served by carrier B (Origin_B), it needs first reallocate from its own origin (Origin_A) to Origin_B. Subsequently, carrier A proceeds to Destination_B before reallocating back to its destination (Destination_A). This reallocation strategy includes three segments, two reallocation routes and one main OD route originally assigned to carrier B. Consequently, carrier B would save cost as it only drives from previous nodes directly to Destination_B, referred to as the ”New route” in Figure 3.2. Conversely, if carrier B carries cargo initially assigned to carrier A, it undertakes two additional reallocation routes. Carrier A, in this case, visits the Destination_A from the node preceding the Origin_A, reducing operational costs.

Two possible modes for collaboration are ”carrier A carry cargo for B”, and ”carrier B carry cargo for A”. The routes after collaborating for both two solutions are also shown in Figure 3.2.

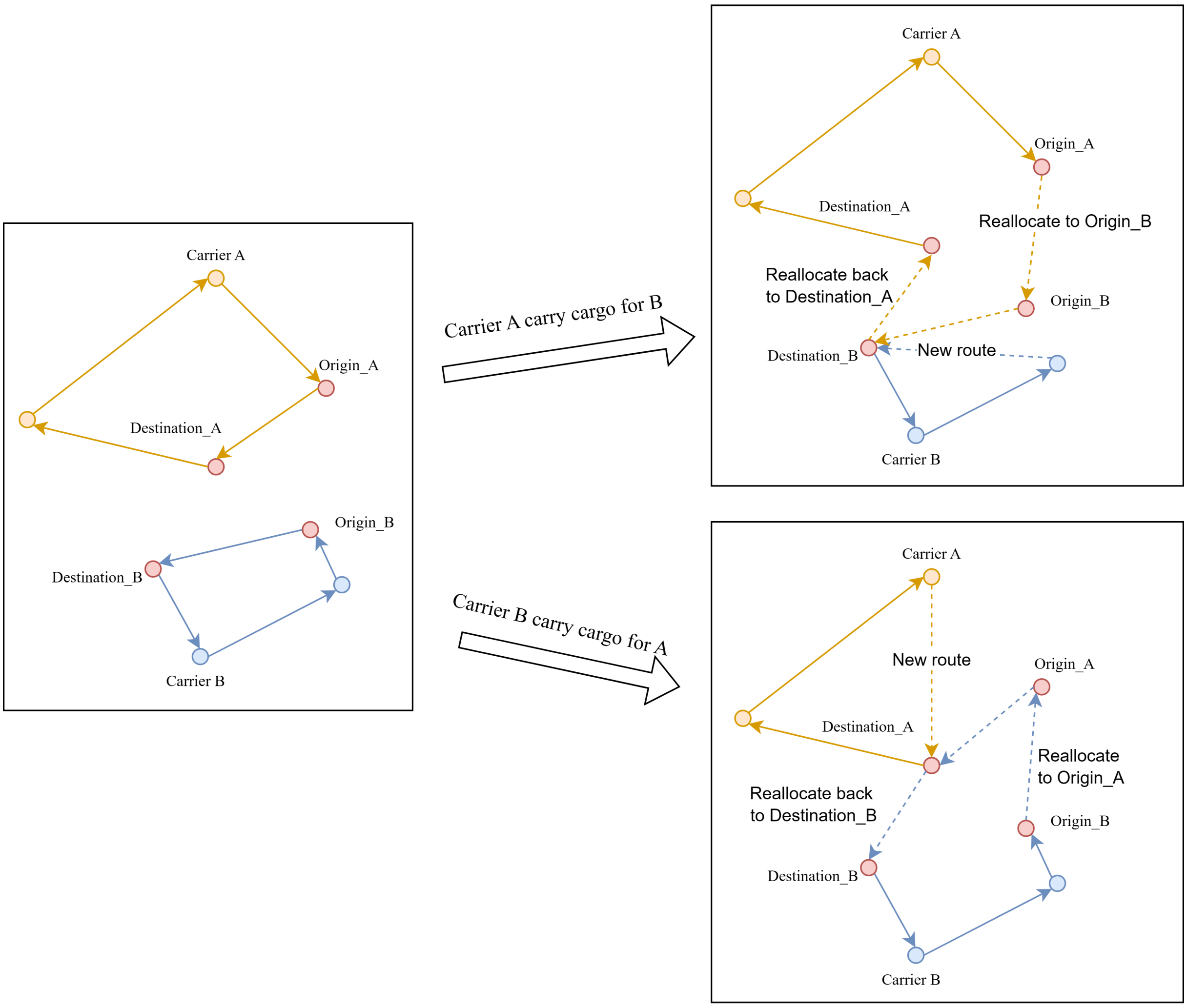


Figure 3.2: Collaboration solutions for OD trips from carrier A and carrier B within a certain range (example 2)

Example 2 further investigates the collaboration solutions based on Example 1 by considering the reallocation between transport tasks. However, some assumptions are still not considered, such as the depots, the compatibility of cargo within a vehicle, and the capacity and weight limitations. These factors are also involved in the horizontal collaboration research of trip pairs from two carriers. The general problem-solving approach, including all these assumptions, will be introduced in the decision tree in Section 3.1.3.

3.1.3 Collaboration decision tree

This section outlines a methodology for identifying four collaboration modes at the trip level using a decision tree. This approach examines the conditions under which two trips from different carriers can potentially collaborate in the context of freight truck transportation under the assumptions introduced in Section 3.1.1. Key variables considered include the routes of the trucks, their capacities, loads, trip weights, and types of cargo. The decision tree provides a logical framework to determine whether trip pairs can collaborate, thereby classifying the collaboration modes.

The decision tree for trip-level collaboration determination is shown in Figure 3.3. The explanation of each part of the decision tree is introduced as follows.

The initial step involves selecting two trips from two different carriers from the freight truck database. The decision tree for collaboration determination focuses on the trip level.

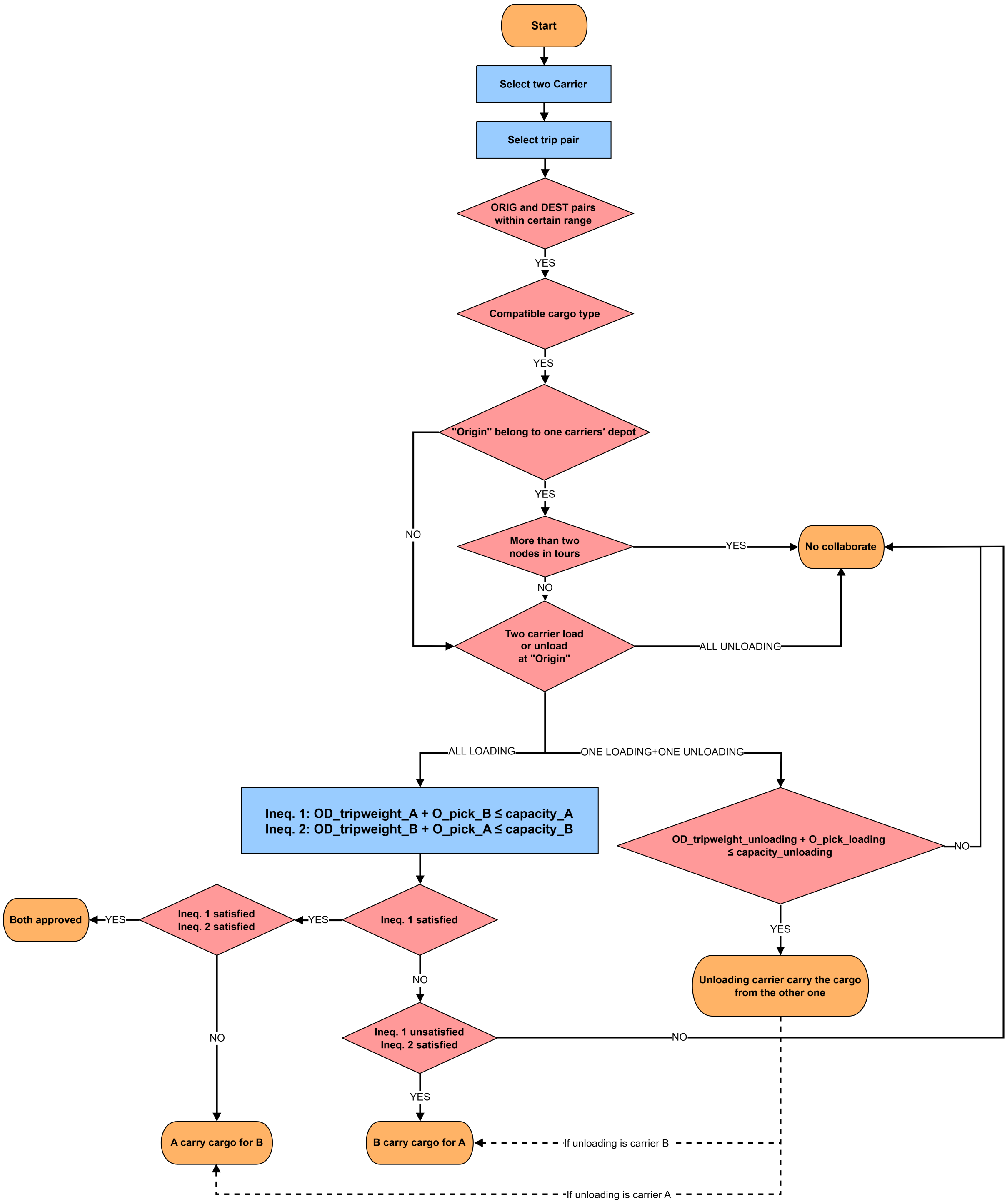


Figure 3.3: Decision tree for trip-level collaboration determination

The analysis aims to distinguish whether the origin and destination (OD) nodes of the trip pairs fall within an acceptable range to facilitate carrier reallocation. Instead of requiring the same OD trip pairs in example 1, the emphasis is on the proximity of these nodes, allowing for routing adjustments (example 2). Specifically, if both carriers' ORIG pair and DEST pair nodes are located within a predetermined range, it suggests a strong potential for collaboration. This approach enables more flexible and efficient coordination without strict adherence to fixed trip routes. The two carrier being determined in the decision tree named as carrier A and carrier B.

Determining the compatible cargo type involves assessing whether different types of cargo can be safely and efficiently stored in the same vehicle. This evaluation ensures that the combined transportation of various cargo types does not compromise the safety or quality of the goods being transported. Trip pairs that do not meet the cargo compatibility criteria are excluded from further consideration.

The next step addresses the scenarios where a carrier′s depot coincides with the ″origin″ in the trip (OD) pairing. Depots serve as critical locations for carriers to prepare cargo and maintain their vehicles, necessitating their inclusion in the route of each tour. In cases where the ″origin″ is a depot for one of the carriers, these cases require detailed examination.

A notable exception arises when a trip involves a depot as the origin and only two nodes are involved along the whole tour. This exception occurs if the cargo can be transported by the other carrier through collaboration. Then the trip involving the depot can be saved and eliminated. Otherwise, collaboration is not feasible under these conditions.

The next step involves analyzing key decision branches within the collaboration process, categorized into three scenarios based on whether the carriers are unloading or loading at the ″Origin″. As explained in Section 3.1.2, only the transport tasks that involve loading at the ″Origin″ are available for sharing. These scenarios are both carriers unloading, both carriers loading, and one carrier unloading while the other is loading.

In the scenario where both carriers are unloading, collaboration is not feasible, as the transport tasks for both carriers are not loading cargo from ″Origin″ to ″Destination″. Conversely, in the mixed scenario of one is loading and the other one is unloading, the carrier that is unloading has the potential to handle the transport task belonging to the other carrier only if that does not exceed its capacity. If this condition is met, collaboration is initiated, with the unloading carrier responsible for the transport task originally belonging to the other carrier. If not, the collaboration attempt fails.

For the scenario where both carriers are loading, two inequalities need to be assessed: Inequality 1 determines if carrier A can carry cargo used to be responsible by carrier B at ″Origin″, while Inequality 2 checks if carrier B can carry the cargo used to be responsible by carrier A. Inequality 1 and Inequality 2 represent two different types of collaboration, as demonstrated in Example 1 and Example 2. Based on the satisfaction of these inequalities, four possible outcomes are listed.

A total of four collaboration modes are proposed: ″A carry cargo for B″ indicates that carrier A and B can collaborate by sharing carrier A's capacity with carrier B; ″B carry cargo for A″ signifies the reverse. ″both approved″ means both carriers can share their capacity with each other, while ″no collaborate″ indicates that carriers A and B cannot collaborate for the chosen trips.

This decision tree outlines how two carrier trip pairs might collaborate based on detailed freight transportation data. The output modes identify the type of collaboration for different trip pairs, setting the stage for further research that will incorporate quantitative analysis to evaluate the performance of these collaborations.

3.2 Collaboration index formulation

The collaboration decision tree identifies four collaboration modes at the trip level. Building on these modes, this section employs a heuristic local optimization approach to evaluate collaboration at the carrier level. This involves analyzing how trip-level collaborations translate to broader carrier-level partnerships. Subsequently, a collaboration index is formulated at the carrier level, which quantifies the collaboration potential between different carriers. This index serves as a comprehensive metric for assessing carrier collaborations in freight transportation.

3.2.1 Heuristic local optimization

The collaboration determination provides collaboration options with four modes (three of which are feasible for collaboration), and an optimization approach needs to be applied to choose the relative best collaboration options between carriers. Given the lack of complete data, particularly cargo origin/destination (O/D) data, a generalized alternative to optimization is considered to develop simple decision rules that are easy to apply in practice. Heuristic local optimization is applied based on the trip-level results of the collaboration identification using the decision tree in Section 3.1.3. The rules of heuristic local optimization are constructed based on the main elements of collaboration found in the literature (Karam et al., 2021) to sort the collaboration options. This method provides a local optimization approach and can be compared to optimization models or data-driven approaches in future studies.

A straightforward method for evaluating collaboration is cost reduction. The predefined cost per kilometre parameters for seven different types of vehicles transform the distance savings (km) achieved through collaboration into monetary cost reduction (€). The cost savings from the collaboration are treated as benefits. These benefits, quantified in monetary terms, directly reflect the savings each carrier achieves through collaboration. The mode "both approved" indicates that the collaboration plans "A carry cargo for B" and "B carry cargo for A" are both feasible for the trip pairs. This mode is unique compared to other modes in that it calculates the benefits for both scenarios and uses their average as the final benefit value. For the "no collaborate" mode, the benefits are set to zero.

The trip-level collaboration determination identifies potential collaboration options for carriers, but the trip pairings are not necessarily unique. For a specific carrier, a single trip may present multiple collaboration opportunities. Instead of solving for the highest global profit through a complex optimization problem, several metrics are defined and applied to sort these trip-level collaboration options, receiving a local optimal for each carrier. These metrics reflect the reference introducing the potential benefits of horizontal collaboration, including lower transport costs, less empty running, less carbon emissions, less traffic congestion, shorter delivery times, higher resource utilization, etc (Karam et al., 2021).

These metrics are considered one by one in the following order:

1. Total Profit: Prioritize collaboration options that offer the highest positive total profit.
2. Frequency of Collaboration: Prefer participant carriers that have more frequency options with Carrier A, as these carriers may indicate more stable and reliable relationships.
3. Tour Savings: Select the option that saves an entire tour for Carrier A, implying fewer vehicles needed and thereby reducing costs.
4. Vehicle Capacity of Participant: Opt for participant vehicles with lower capacity since smaller vehicles generally incur lower costs, enhancing flexibility and efficient resource use in the collaboration.
5. Departure Time Differences: Choose options with the smallest differences in departure times with trip in carrier A to minimize waiting times and potential scheduling conflicts.

Viable options are then selected in order, ensuring that each trip from a carrier is paired only once with the most suitable trip from a participating carrier. This method ensures that each trip pair appears uniquely in the selection process, preventing redundancy and repeated pairings in decision-making and analysis.

Based on the five metrics building for the sorting metrics, and the unique selection, each carrier identifies its own optimal collaborative plan. Although these plans give suggestions for each carrier, showing the optimal collaborative plan from their individual perspective, the plans do not lead to a globally optimal solution. This is because conflicts can arise between the plans of different carriers. For instance, a trip from carrier A might be best to collaborate with a trip from carrier B to optimize carrier A's benefits. However, from carrier B's perspective, the same trip might achieve greater profitability if it collaborates with a trip from carrier C.

A simple method to eliminate the conflict options between each carrier is to find the same options between carriers. This is achieved by finding the intersection of the local optimal options across all involved carriers. The solutions that appear in this intersection are considered for the final decision, as they represent the options that all carriers agree upon. This "Intersection Selection" aims to avoid conflict among carriers involved. While it may not yield a globally optimal solution, it ensures a higher degree of mutual satisfaction and feasibility for implementation. The final plans after eliminating the conflict for each carrier show the feasible collaborative solutions.

After summing the benefits of the final plan up by the carrier, a matrix, referred to as the “Carrier Benefits Matrix”, displays the benefits received from the feasible collaborative solutions for each carrier after resolving conflicts. It quantifies summing up the cost reduction (benefits) from the trip level to the carrier level. This matrix highlights the total benefits realized between carriers, showing the total economic gain from the collaboration for the participants.

Since this matrix shows the economic gains achieved through collaboration, it effectively serves as an index of collaboration effectiveness. Consequently, the 'Carrier Benefits Matrix' can also be named the 'Carrier Collaboration Index Matrix', as it provides a clear indication of the level of successful financial collaboration between the carriers.

4 Data overview

This chapter provides a comprehensive overview of the data utilized in this research. The first section, Data Description, introduces the dataset used and highlights the key attributes relevant to the study. In the second section, Data Exploration, various analyses are conducted to discover insights from the pre-processed data. This includes visualizing tours, examining the distribution of trip distances, analyzing the load rate distribution, and presenting an example of the same trip pairs for two carriers. These explorations aim to provide a deeper understanding of the data and its implications for the research.

4.1 Data description

The dataset utilized in this research comprises extensive freight transportation data, capturing the operational trajectories of hundreds of carriers engaged in pickup and delivery services. The synthesized data was generated by MASS-GT based on freight tour data collection by the Central Bureau of Statistics Netherlands (CBS) (de Bok & Tavasszy, 2018; Mohammed et al., 2023). However, the customer data for each carrier, including detailed origin and destination information for cargo, is not available. Only the origin-destination (OD) data are provided and production-consumption (PC) data are not available. Each row within the dataset corresponds to a specific trip undertaken by a carrier, with detailed information spanning 17 attributes.

Key among these attributes are:

- *CARRIER_ID*: This field assigns a unique identifier to each carrier. Within the dataset, a single carrier may undertake multiple tours. A total of 316 carriers are included.
- *TOUR_ID*: This identifier denotes specific tours undertaken by a carrier. For instance, the notation "0_0" in the data frame indicates that the corresponding row refers to the first tour (Tour 0) of the first carrier (Carrier 0).
- *TRIP_ID*: Similar to the TOUR_ID, this identifier marks individual trips within a tour. An entry marked as "0_0_0" signifies the first trip (Trip 0) of the first tour (Tour 0) of the first carrier (Carrier 0).
- *ORIG & DEST*: These identifiers mark the origin and destination of each trip.
- *X_ORIG, X_DEST, Y_ORIG, Y_DEST*: These are coordinates of origin and destination.

- *VEHTYPE*: This refers to the types of vehicles, and different vehicles have corresponding different capacities. A total of seven types of vehicles are included in the dataset.
- *N_SHIP*: This refers to the number of commodities that are inside the truck while travelling between this origin and destination pair.
- *DC_ID*: This identifier shows the position where the carrier's depot belongs to.
- *TOUR_WEIGHT*: The total weight being loaded in the vehicle along the tour.
- *TRIP_WEIGHT*: The weight being transported in the vehicle in the corresponding trip.
- *TRIP_DEPTime*: The trip departure time at the origin.
- *TRIP_ARRTIME*: The trip arrival time at the destination.
- *LOG_SEG*: This refers to the logistic segment, indicating the types of cargo. Nine different types of cargo are listed.

These attributes facilitate a detailed view of freight transport operations at the trip level across various carriers. Additionally, the dataset enables the computation of other operational attributes such as load rate and pickup demand. After filtering out the missing and error data, this research utilizes a prepossessed dataset comprising 95,737 records across 315 carriers.

In the dataset, the "LOG_SEG" attribute categorizes the types of cargo transported along the trips, listing a total of nine different cargo types. The concept of compatible cargo type is crucial, as it determines whether different types of cargo can be safely stored within the same vehicle. To facilitate this, a compatibility matrix, shown in Table 4.1, is assumed and utilized to identify which cargo types can be transported together in the same vehicle. This matrix serves as a guideline for ensuring safe cargo transportation by preventing incompatible cargo combinations. Value one means available and value zero means unacceptable. The matrix is assumed and defined by common sense. For instance, combining food with dangerous materials in the same vehicle is not acceptable. Trip pairs that do not meet the cargo compatibility criteria are excluded from collaborative consideration.

Table 4.1: Compatible cargo type matrix

	Food (general cargo)	Miscellaneous (general cargo)	Temperature controlled	Facility logistics	Construction logistics	Waste	Parcel (consolidated flows)	Dangerous	Parcel (deliveries)
Food (general cargo)	1	1	1	0	0	0	1	0	1
Miscellaneous (general cargo)	1	1	1	0	0	0	1	0	1
Temperature controlled	1	1	1	0	0	0	1	0	1
Facility logistics	0	0	0	1	1	0	0	0	0
Construction logistics	0	0	0	1	1	0	0	0	0
Waste	0	0	0	0	0	1	0	0	0
Parcel (consolidated flows)	1	1	1	0	0	0	1	0	1
Dangerous	0	0	0	0	0	0	0	1	0
Parcel (deliveries)	1	1	1	0	0	0	1	0	1

The straightforward method to avoid complex calculations involves using a predefined cost per kilometre for the seven different vehicle types included in MassGT (de Bok & Tavasszy, 2018; Mohammed et al., 2023). The unit of this cost is expressed in €/km. All parameters are summarized in Table 4.2.

Table 4.2: Cost per kilometre for different vehicle types

VEHTYPE	CostPerKm
Truck_Small	0.1954
Truck_Medium	0.3404
Truck_Large	0.5724
TruckTrailer_Small	0.2244
TruckTrailer_Large	0.5144
TractorTrailer	0.4274
SpecialVehicle	0.4274

4.2 Data exploration

In this section, a journey is undertaken to discover insights from the pre-processed dataset described in the previous section.

Each row of the data is based on the trip level, the distance for each trip can also be explored. The histogram can distribute the data into bins, identifying the variability of trip distances within the data. The determination of bin size would influence the visualization to reflect the trip distance distribution. For this research, the bin size was determined using Sturges' rule. Sturges' rule suggests setting the number of bins to $1 + \log_2(N)$, where N is the number of data points which is 95,737. By employing this approach, the bin size is set to 18. Figure 4.1 shows the trip distance distribution of the data. The figure indicates frequent trips within the initial bins, particularly in the 0-25,121 meter range. The first bin includes more than 23,000 trips, suggesting that most of the journeys are relatively short. As distances increase, a marked decline in the number of trips is observed. Extremely long journeys, such as those about 400,000 meters are infrequent. This could reflect specific logistic requirements for long distances within the data.

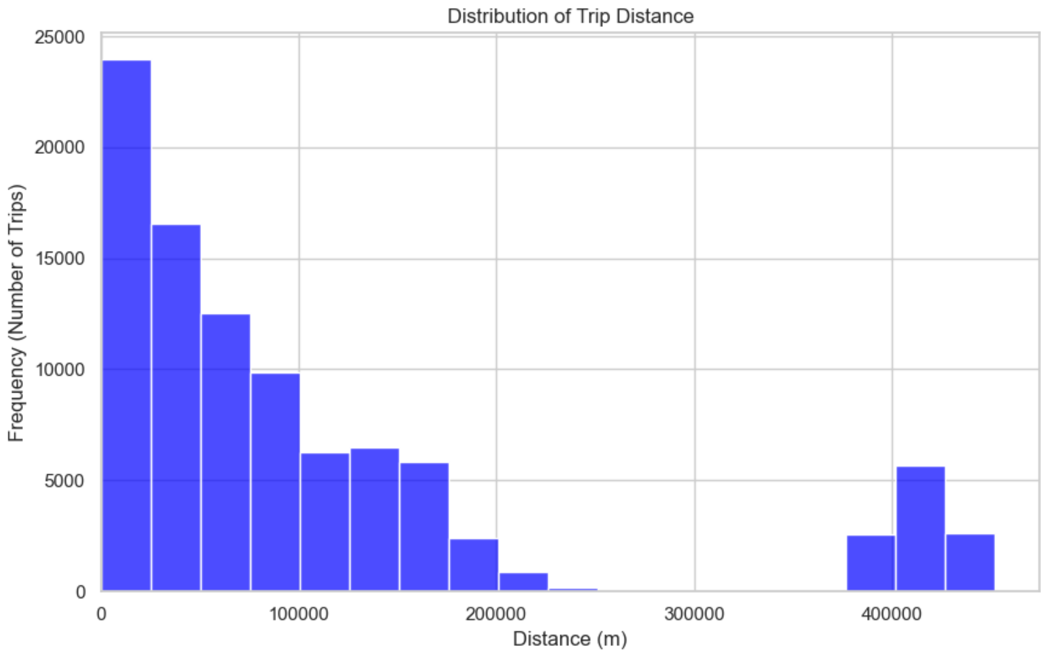


Figure 4.1: Distribution of trip distance

The trip load rate is an indicator of the efficiency of freight transportation. A low load rate for carriers indicates inefficiency and potential capacity for sharing with partners in horizontal collaboration.

Figure 4.2 illustrates the distribution of the proportion of trips where the load rate is below 50% for various carriers. The horizontal axis represents the proportion of such trips, while the vertical axis shows the number of carriers. The chart highlights the potential for collaboration to optimize load utilization.

The figure reveals significant variability in the proportion of low-load-rate trip journeys, with all carriers within the data exhibiting low-load trips above 70%. Most carriers having a low load rate trips proportion between 0.775 and 0.85. This indicates that for the majority of carriers, about 77.5% to 85% of their trips have a load rate below 50%. This high proportion of low load rate trips highlights the current low utilization of available capacity and suggests significant potential for optimization and collaboration.

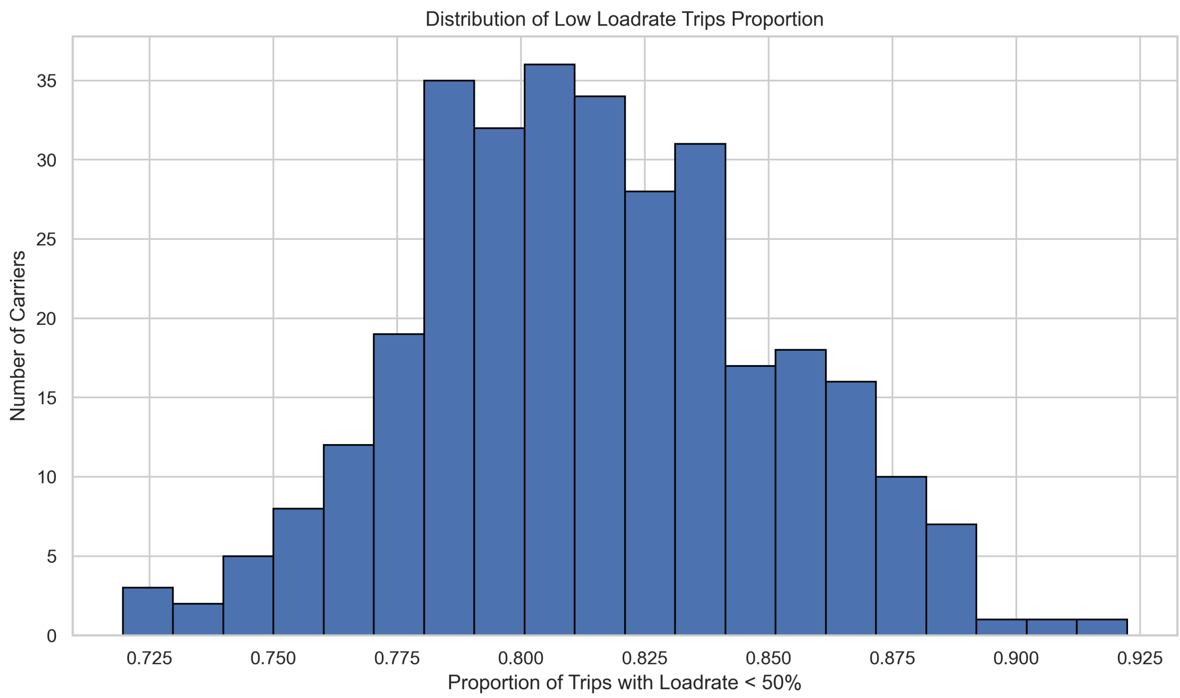


Figure 4.2: Distribution of the proportion of trips with load rate below 50% among carriers

If reallocation is not considered, the same origin-destination (OD) trips between different carriers reveal a close relationship and potential collaborative opportunities, as introduced in Chapter 3. Figure 4.3 illustrates the frequency of trips for different carriers between origin and destination pairs. Each line represents a trip path taken by multiple carriers. Each subplot represents a different range of OD counts: 1 to 50, 51 to 100, 101 to 200, and 201 and above. The colour gradient, ranging from light blue to dark blue, represents the frequency of different carriers recording the same origin-destination (OD) trip. Darker shades indicate higher trip frequencies, highlighting routes with more frequent travel by different carriers. This visualization shows the potential for collaboration on these identical high-frequency OD trips.

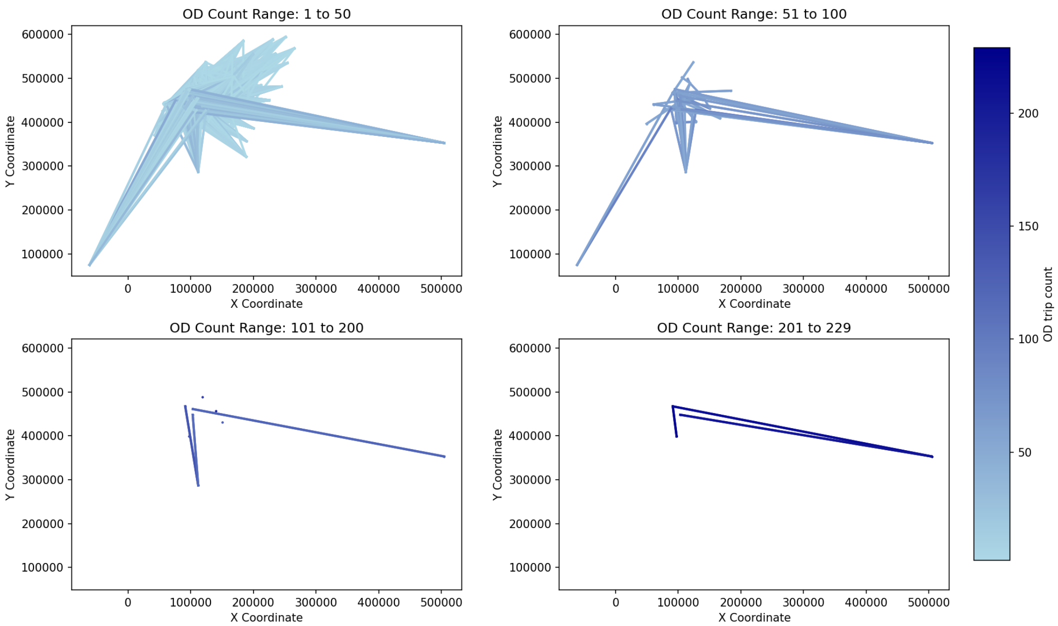


Figure 4.3: Frequency of same OD trips by different carriers

The visualization presented in Figure 4.3 offers a comprehensive view of the frequency of the same OD trips. Here a simple same OD trip example is given, Figure 4.4 focuses on the trip level, illustrating two tours by carrier 65 and carrier 232. These tours follow the same trip route from node 268 to node 3840. Visiting the same nodes is an insight into horizontal collaboration, such as sharing capacity and enhancing transportation efficiency. For instance, one carrier could handle the transport task for a shared trip, allowing the other carrier to optimize operations and improve efficiency. This example highlights carriers visiting the same OD trip, but generally, trip pairs within acceptable proximity also offer potential for collaboration. Further details on determining collaboration opportunities will be explored in the next chapter.

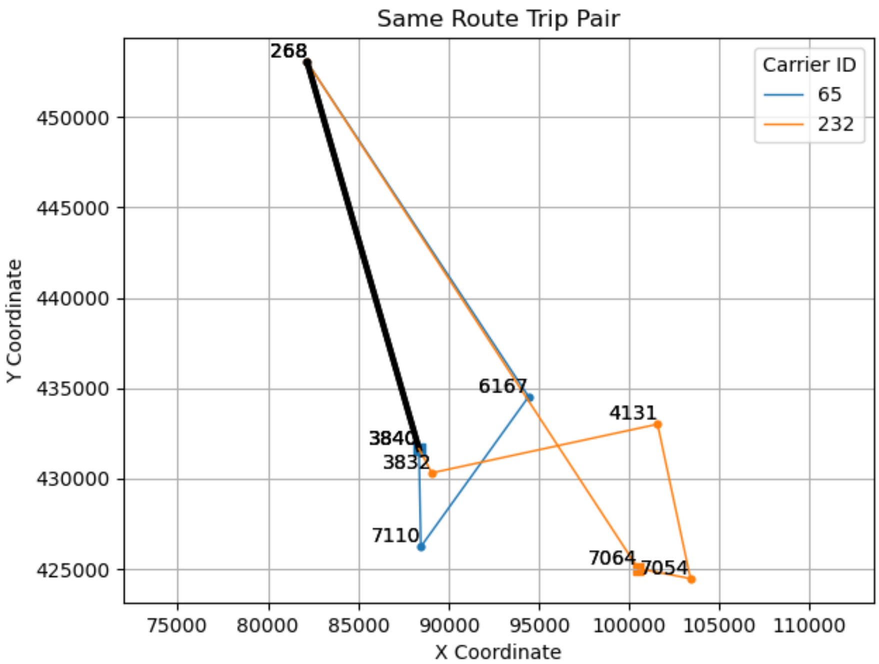


Figure 4.4: Visualization trip pair routes from carrier 65 and carrier 232

5 Implementation

This chapter constructs several scenarios based on the parameters and variables included in the methodology to facilitate a deeper examination of the results. These scenarios serve as frameworks for testing various understanding the implications of the collaboration index from different perspectives. By applying these scenarios, the result analysis can discover insights of the collaboration index.

The limit range for allocation, discussed in Section 3.1.3, allows for routing adjustments between trips. This predefined acceptable range for allocation influences the number of collaboration opportunities available, which in turn affects the collaboration index between carriers. By adjusting this range, the impact on the collaboration index is investigated, providing insights into how flexible routing parameters can enhance or limit the potential for effective collaboration between carriers.

The heuristic local optimization provides a simple approach to aggregate trip-level collaboration modes into the carrier-level index. After the collaboration options are sorted according to the five metrics presented in Section 3.2.1, two additional variables are considered before the selection procedure: random selection rate and cost reduction threshold.

The random selection rate is employed to illustrate that carriers might not consider all available options when exploring collaboration. Instead, there is a probability that they will choose to collaborate at a specific rate. Considering this rate of selection introduces additional randomness into the computation of benefits. Moreover, it helps in demonstrating how the benefits fluctuate based on variations in the selection rate.

The threshold is employed to filter out less impact collaboration options, allowing carriers to concentrate their resources and efforts on collaborations that offer significant cost savings. This cost reduction threshold is used to assess the viability of potential collaborative options. The cost reduction rate is defined as the percentage of cost savings relative to the total cost of the impacted routes. Specifically, a collaboration option is only considered viable if the cost savings achieved surpass a predefined percentage threshold.

Table 5.1 presents 54 scenarios generated through three variables: limit range, random selection rate, and cost reduction thresholds. The unique combinations derived from varying levels of each variable—limit range (0, 500, 1000 km), random selection rate (100%, 80%, 60%), and cost reduction thresholds (0%, 10%, 20%, 30%, 40%, 50%). The exploration of these scenarios helps investigate how each variable impacts the performance and cost-saving potential of collaboration. This analysis also contributes to the development of the collaboration index. The detailed analysis is presented in Chapter 6.

Table 5.1: Scenarios for limit range, random selection, and cost reduction threshold

Limit Range (m)	Random Selection Rate (%)	Cost Reduction Threshold (%)
0	100	0
0	100	10
0	100	20
0	100	30
0	100	40
0	100	50
0	80	0
0	80	10
0	80	20
0	80	30
0	80	40
0	80	50
0	60	0
0	60	10
0	60	20
0	60	30
0	60	40
0	60	50
500	100	0
500	100	10
500	100	20
500	100	30
500	100	40
500	100	50
500	80	0
500	80	10
500	80	20
500	80	30
500	80	40
500	80	50
500	60	0
500	60	10
500	60	20
500	60	30
500	60	40
500	60	50
1000	100	0
1000	100	10

1000	100	20
1000	100	30
1000	100	40
1000	100	50
1000	80	0
1000	80	10
1000	80	20
1000	80	30
1000	80	40
1000	80	50
1000	60	0
1000	60	10
1000	60	20
1000	60	30
1000	60	40
1000	60	50

The 'Carrier Collaboration Index Matrix' is the output of this research, providing a direct relationship between carriers, indicating the potential collaboration opportunities and the level of benefits that may shared with the participants.

The 'Carrier Collaboration Index Matrix' is the key output of this research. It provides a direct relationship between carriers, indicating potential collaboration opportunities and the level of benefits that would be shared among the participants. This matrix serves as a tool for carriers to identify and evaluate possible partnerships, enhancing their decisions regarding collaborative strategies and optimizing overall operational efficiency. The detailed application for the collaboration index is introduced in Section 6.2.

6 Result analysis

In this chapter, the scenarios developed in Chapter 5 are applied. In the first section, the results, including the total benefits and cost-saving rate under different scenarios, are analyzed through sensitivity analysis. This analysis provides insights into the effectiveness and robustness of the proposed strategies under various conditions.

In the second section, the collaboration index obtained from this research is presented. Additionally, several analyses of the index are conducted, offering relevant advice on horizontal collaboration strategies based on the findings. An in-depth analysis based on the collaboration index for interpretation is followed. This comprehensive evaluation aims to enhance the understanding and application of the collaboration index in real-world transportation logistics.

6.1 Sensitivity analysis

This section presents the effect of limit range, random selection rate and cost reduction threshold on the performance of total benefits of the final collaboration plan and total cost saving rate. All of them are based on the carrier level.

The total benefits of the final collaboration plan exclude the option conflict as introduced in Section 3.2.1. The value of total benefits shows the total amount of benefits that could gained from the collaboration for the data under the specific scenarios. The total cost-saving rate is calculated by the following formula.

$$\text{Total cost saving rate (\%)} = \left(\frac{\text{Total\_Cost\_saving}}{\text{Total\_Cost}}\right) \times 100 \tag{6.1}$$

The $Total\_Cost$ refer to the total cost for all the trips involved in the data without collaboration. The total cost-saving rate shows the percentage of benefits that account for the total cost, which emphasizes the effectiveness of collaboration.

Table 6.1 shows the result of total benefits and cost savings varies different limit range (0m, 500m, 1000m), random selection rate (100%, 80%, 60%) and cost reduction threshold (0%, 10%, 20%, 30%, 40%, 50%).

Table 6.1: The total benefits and cost savings under different limit range (0m, 500m, 1000m) random selection rate (100%, 80%, 60%) and cost reduction threshold (0%, 10%, 20%, 30%, 40%, 50%)

Limit Range (m)	Random Selection Rate (%)	Cost Reduction Threshold (%)	Total Benefits (exclude option conflict) (€)	Total Cost Saving Rate (exclude option conflict) (%)
0	100	0	706178.83	22.77
0	100	10	705278.58	22.74
0	100	20	685293.4	22.09
0	100	30	638525.03	20.59
0	100	40	350187.36	11.29
0	100	50	210424.81	6.78
0	80	0	491190.14	15.84
0	80	10	490522.38	15.82
0	80	20	475545.97	15.33
0	80	30	441999.86	14.25
0	80	40	244933.38	7.90
0	80	50	148159.72	4.78
0	60	0	330276.61	10.65
0	60	10	329721.15	10.63
0	60	20	319489.2	10.30
0	60	30	296876	9.57
0	60	40	161735.03	5.21
0	60	50	94947.98	3.06
500	100	0	808220.39	26.06
500	100	10	807197.72	26.03
500	100	20	785400.67	25.32
500	100	30	735835.81	23.72
500	100	40	406416.93	13.10
500	100	50	245173.15	7.90
500	80	0	565005.05	18.22
500	80	10	564280.96	18.19
500	80	20	546593.74	17.62
500	80	30	511483.31	16.49
500	80	40	274956.06	8.86
500	80	50	167095.76	5.39
500	60	0	389689.69	12.56
500	60	10	389210.29	12.55
500	60	20	375858.93	12.12
500	60	30	352105.65	11.35
500	60	40	190894.55	6.15
500	60	50	115234.17	3.72

1000	100	0	907079.36	29.25
1000	100	10	905636.42	29.20
1000	100	20	880489.92	28.39
1000	100	30	826381.57	26.64
1000	100	40	457465.95	14.75
1000	100	50	282382.60	9.10
1000	80	0	625723.88	20.17
1000	80	10	624681.91	20.14
1000	80	20	605715.39	19.53
1000	80	30	566466.92	18.26
1000	80	40	310800.22	10.02
1000	80	50	195216.18	6.29
1000	60	0	443101.57	14.29
1000	60	10	442442.79	14.27
1000	60	20	428681.85	13.82
1000	60	30	401124.78	12.93
1000	60	40	218271.42	7.04
1000	60	50	136574.88	4.40

6.1.1 Effect of the Limit Range

The effect of the Limit Range on the total cost-saving rate is illustrated in Figure 6.1. Each subplot corresponds to a different Random Selection Rate (100%, 80%, 60%), and the lines represent different cost reduction thresholds. It is evident from the figure that as the Limit Range increases from 0 to 1000, the total cost-saving rate generally shows an upward trend. This suggests that increasing the limit range expands the decision-making space, allowing for more reallocation considerations and including more potential trips in the determination process. Consequently, this enhances benefits and cost savings, as a larger limit range facilitates the identification and optimization of more collaborative opportunities.

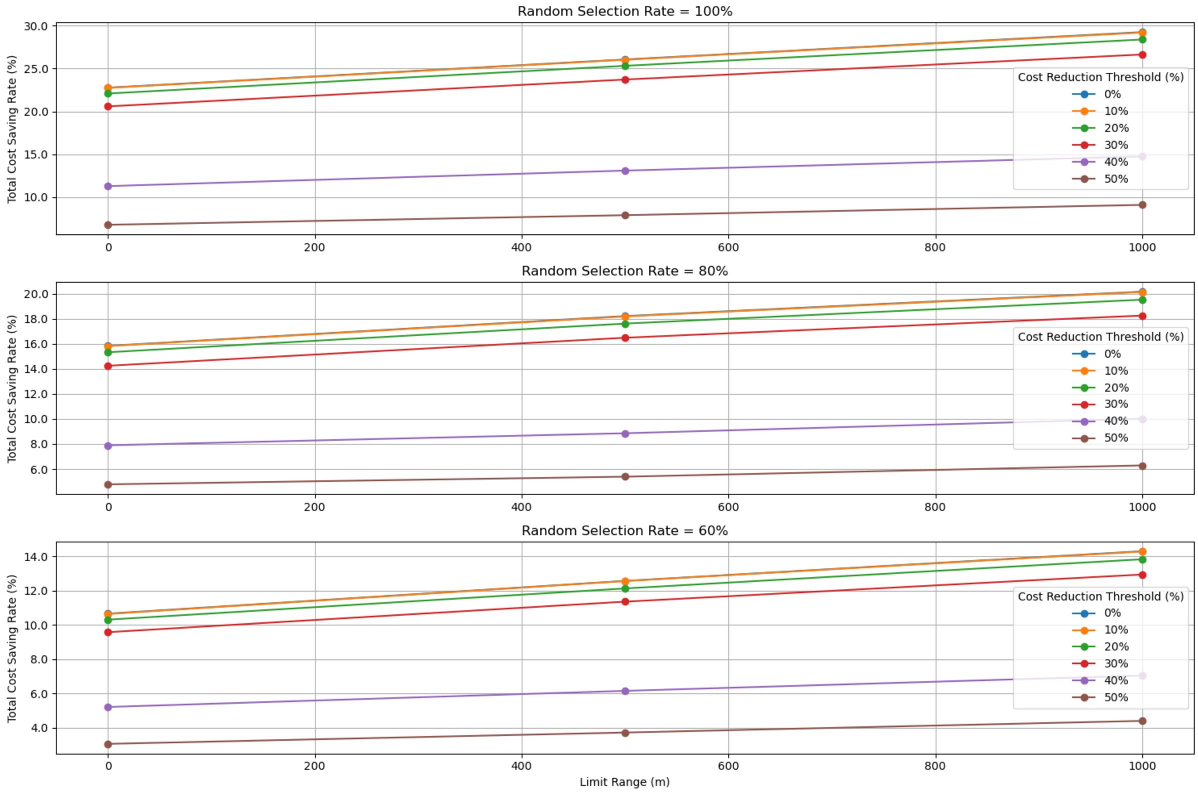


Figure 6.1: The relationship between Total Cost Saving Rate and Limit Range across different Random Selection Rate and Cost Reduction Thresholds

6.1.2 Effect of the Random Selection Rate

Figure 6.2 illustrates the total cost-saving rate as a function of the random selection rate (%) for various cost reduction thresholds. Each subplot corresponds to a different Limit Range (0m, 500m, 1000m), showcasing how cost savings fluctuate across different selection rates and thresholds.

The Random Selection Rate is used to show the carriers may not consider all collaboration options for collaboration. The randomness added to the model also helps to indicate the fluctuation of benefits. When the Random Selection Rate decreases (from 100% to 60%), irrespective of the Limit Range, the total cost-saving rate tends to decrease. This indicates that a higher random selection rate improves project benefits and cost efficiency because greater randomness offers more potential favorable collaboration combinations.

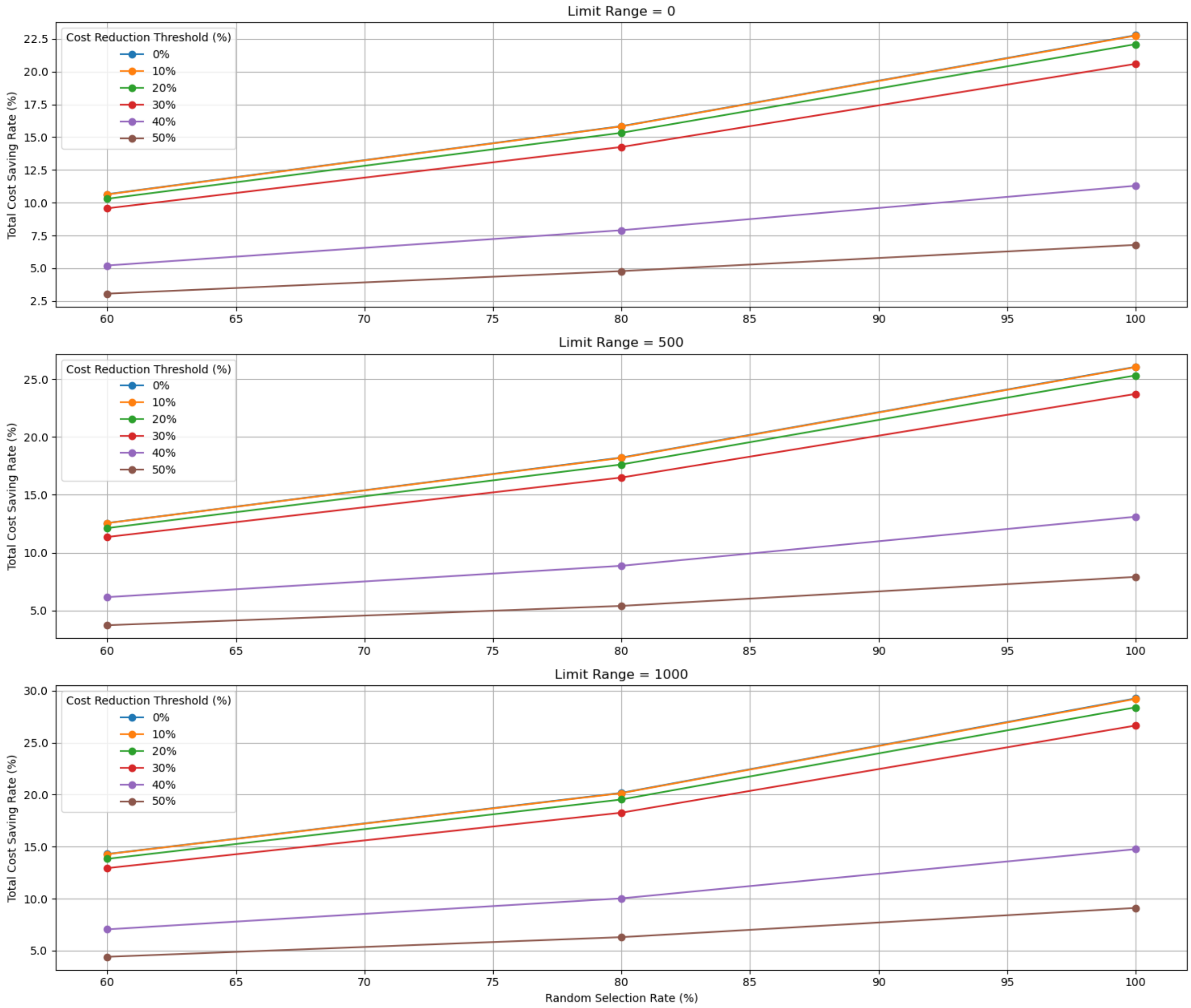


Figure 6.2: The relationship between Total Cost Saving Rate and Random Selection Rate across different Limit Ranges and Cost Reduction Thresholds

6.1.3 Effect of the Cost Reduction Threshold

The Cost Reduction Threshold is employed to enable carriers to concentrate their resources on collaborations that yield high cost savings. However, since the cost calculations in this research consider only distance and do not account for other realistic expenses such as wages and storage costs, collaborations with low cost savings might result in negative revenue in actual practice. Implementing this threshold helps to mitigate such risks by ensuring that only those collaborations with significant cost-saving potential are pursued, thereby maximizing the likelihood of positive financial outcomes for the carriers.

Figure 6.3 shows the effect of the Cost Reduction Threshold on the total cost-saving rate. As the Cost Reduction Threshold increases, both the total benefits and the total cost-saving rate generally decline. This is because higher Cost Reduction Thresholds eliminate collaboration options with lower benefits, thus reducing the overall benefits. The slope of the decrease continues to steepen, with a significant drop observed when the Cost Reduction Threshold exceeds 30%. This indicates that beyond this threshold, the exclusion of potential collaborations with relatively low cost savings markedly impacts the total cost-saving rate, highlighting the trade-off between stricter criteria and overall savings.

For example, when the Limit Range is set at 1000 meters and the Random Selection Rate at 100%, the slopes are recorded as (-0.005, -0.081, -0.174, -1.190, -0.564). These values represent the slopes, and their progression illustrates a gradual transition from a small negative value to increasingly larger negative values. The trend in these slopes indicates an initially slow decline that accelerates rapidly before slightly decelerating, yet it still maintains a rapid downward trajectory. The significant drop observed between the 30% and 40% thresholds may indicate that many of the collaboration decisions occur within this range for the dataset explored in this research. One contributing factor to the observed decline is that within the 30 to 40 Cost Reduction Threshold, many collaboration plans involving one of the carriers only include two nodes throughout the entire tour. Specifically, the entire transport task of the tour could be managed by the partner carrier if collaboration occurs. Although reallocation led to increased costs due to detours, the collaboration plan still yielded positive benefits. This arrangement facilitated the sharing of reduced costs generated through collaboration, thereby enabling profit sharing between the carriers.

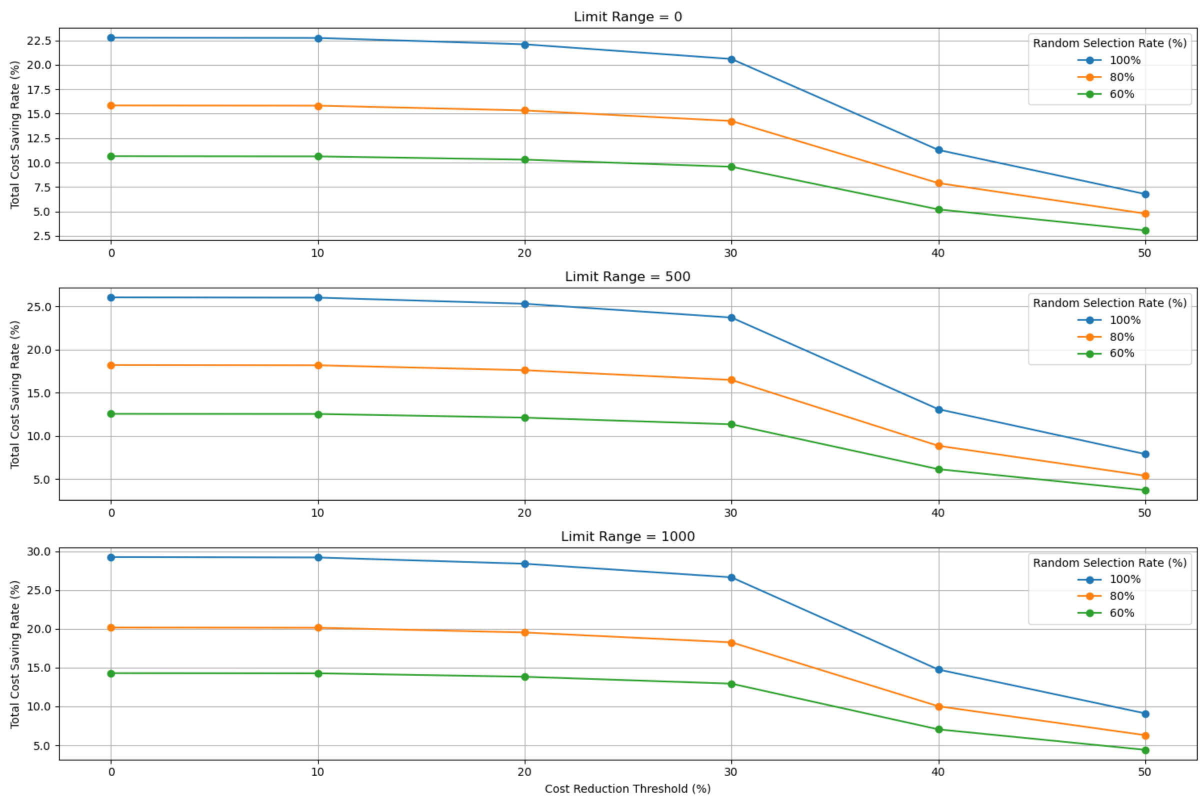


Figure 6.3: The relationship between Total Cost Saving Rate and Cost Reduction Thresholds across different Limit Ranges and Random Selection Rate

6.2 Collaboration index analysis

Section 6.1 investigated the effects of three variables on the benefits and cost-saving rate of the collaboration. Based on these findings, a Limit Range of 1000m, a Random Selection Rate of 100%, and a Cost Reduction Threshold of 20% are applied in this section. These values allow for reallocation between trips, with carriers following all collaboration decision suggestions and focusing on high cost-saving collaboration options by following a reasonable threshold. The output of the collaboration index, derived from these parameters, will be analyzed in this section, providing insights into the horizontal collaboration strategies and the suggestions provided to the carriers.

The data for this research includes a total of 316 carriers. The resulting "Carrier Collaboration Index Matrix" has dimensions of 316 by 316. A part of the matrix is shown in Table 6.2. This example illustrates how the collaboration index can be utilized in freight transportation and highlights the suggestions derived from the collaboration index. By examining this matrix, carriers can identify their optimal partners for collaboration, focusing on those with high potential benefits. This enables carriers to make decisions and implement collaboration solutions that increase efficiency and save costs.

Table 6.2: Collaboration index for carriers with IDs ranging from 20 to 30

Carrier_ID	20	21	22	23	24	25	26	27	28	29	30
20	0.00	375.06	245.42	208.17	488.05	0.00	0.00	0.00	0.00	135.50	0.00
21	375.06	0.00	275.32	1277.09	360.55	0.00	0.00	0.00	0.00	0.00	86.74
22	245.42	275.32	0.00	994.52	55.45	0.00	0.00	0.00	0.00	114.41	42.59
23	208.17	1277.09	994.52	0.00	2104.28	0.00	0.00	0.00	0.00	202.54	169.74
24	488.05	360.55	55.45	2104.28	0.00	0.00	0.00	0.00	0.00	722.29	222.44
25	0.00	0.00	0.00	0.00	0.00	0.00	407.03	461.56	593.79	18.46	0.00
26	0.00	0.00	0.00	0.00	0.00	407.03	0.00	1457.30	607.79	0.00	318.07
27	0.00	0.00	0.00	0.00	0.00	461.56	1457.30	0.00	1634.04	129.72	12.18
28	0.00	0.00	0.00	0.00	0.00	593.79	607.79	1634.04	0.00	433.96	193.20
29	135.50	0.00	114.41	202.54	722.29	18.46	0.00	129.72	433.96	0.00	495.49
30	0.00	86.74	42.59	169.74	222.44	0.00	318.07	12.18	193.20	495.49	0.00

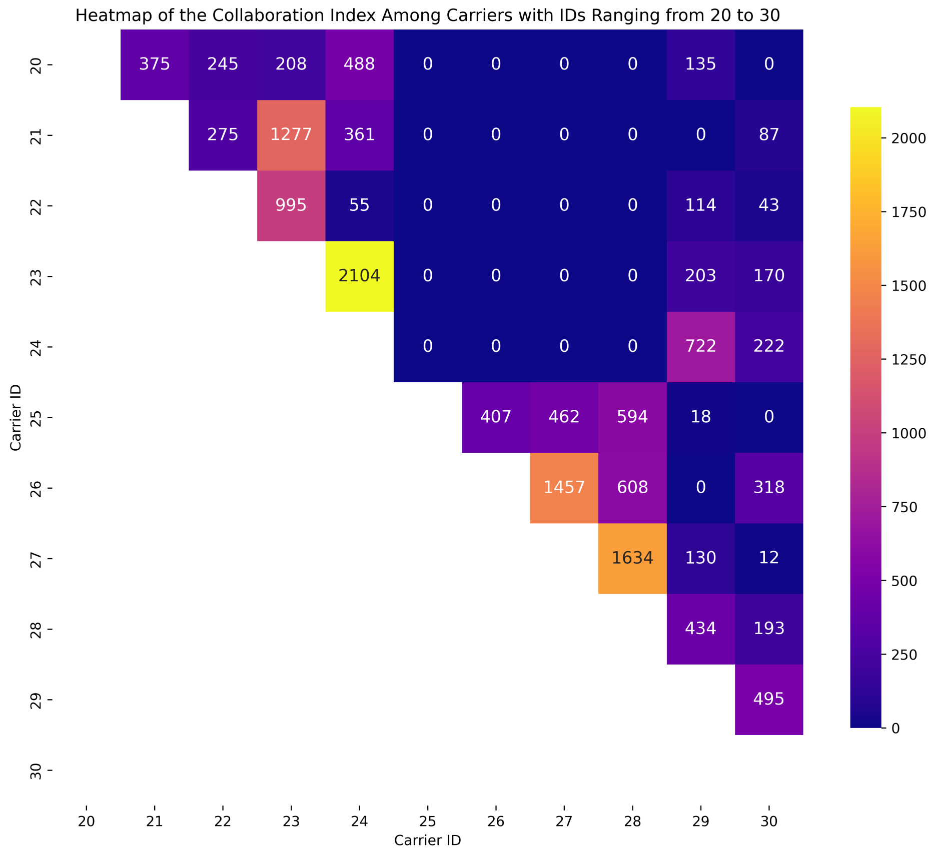


Figure 6.4: Heatmap of the collaboration index among carriers with IDs ranging from 20 to 30

The collaboration index is shown in Table 6.2 includes carriers with IDs ranging from 20 to 30. For a carrier willing to collaborate, the collaboration index reveals potential opportunities. The row corresponding to the carrier's ID provides this information. For example, if carrier 25 wants to check potential partners for collaboration, the values in row 25 of the Collaboration Index Matrix will be provided. The value of the collaboration index indicates the amount of benefits (€) that can be shared when collaborating with the corresponding carrier ID in the column. A higher value of the collaboration index signifies a greater potential for profit sharing in the partnership, guiding carriers towards the most beneficial collaborations.

The heat map displayed in Figure 6.4 illustrates the collaboration index (CI) among carriers, highlighting the varying potential for collaboration between different carriers. The colour scale transitions smoothly from dark purple, representing lower CI values, to bright yellow, indicating higher CI values. Brighter colours denote the higher potential for collaboration and greater benefits shared following its application.

On behalf of carrier 25, the potential shared benefits with carriers 26, 27, and 28 are substantial (407, 462, 594), while the benefits generated from collaboration with carrier 29 are relatively low (18). Although the Cost Reduction Threshold is set at 20%, which eliminates options with slight benefits from the solutions, carrier 25 still needs to assess whether collaboration with carrier 29 would be efficient enough to bring positive benefits in practice. Additionally, detailed decisions for collaboration with different partners are accessible to the carrier through the Collaborative Decision Support System. This system provides comprehensive information on trip-level details, modes of collaboration, collaborative planned routes, and the loading or unloading cargo solutions for vehicles, enabling carrier 25 to make well-informed decisions regarding potential partnerships.

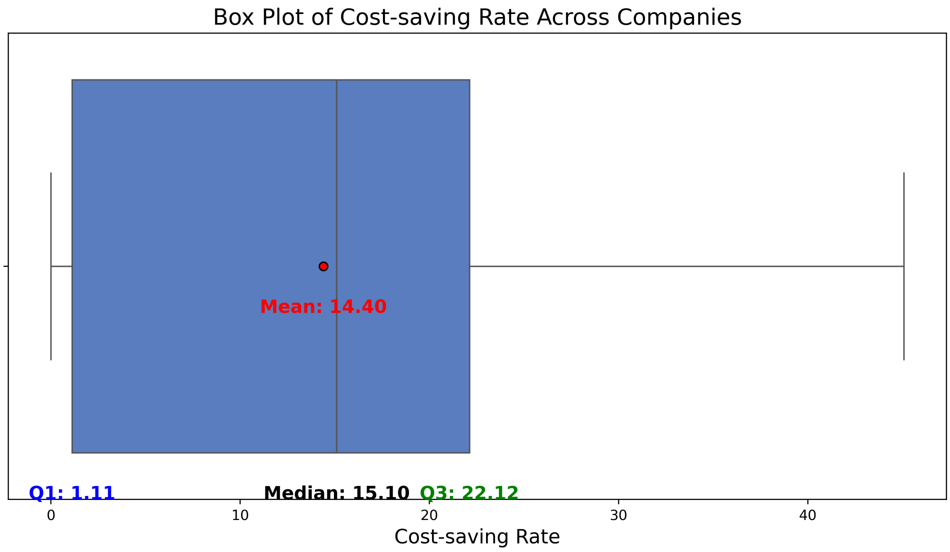


Figure 6.5: Distribution of cost-saving rates across companies from horizontal collaboration

The collaboration index indicates the amount of benefits that can be shared between partners when they collaborate. The cost-saving rate for each participant in the collaboration is also influenced by the total cost across the entire tours in the data. The cost-saving rate helps participants better understand the expected cost reduction impact of horizontal collaboration if they collaborate based on the collaboration plan suggestions.

The box plot shown in Figure 6.5 illustrates the distribution of the cost-saving rate across all companies engaged in horizontal collaboration. The first quartile (Q1), which indicates that the lowest 25% of data points fall below this value, is 1.11%. The median (Q2) is 15.10%, providing a central point that separates the lower and upper halves of the data. The mean cost-saving rate is 14.40%, closely aligning with the median, suggesting a relatively symmetrical distribution of savings. The third quartile (Q3), where the lowest 75% of data points fall below this value, is 22.12%. These quartiles highlight the variability in cost-saving rates among companies. On average, participants in the horizontal collaboration benefit from a 14.4% cost-saving rate, demonstrating the potential advantages of such partnerships.

6.2.1 Collaboration index interpretation

While this research develops a two-step method for examining and identifying horizontal collaboration using the collaboration index, an in-depth analysis of the collaboration index involves explaining and interpreting how various attributes influence its value. In this section, the impact of geographical location and vehicle types is investigated.

Geographical relation

In Section 3.1.3, geographical location is used to determine the distance between two trips and whether the origin and destination nodes are within a certain range. Here, a higher-level analysis focuses on the overall node locations, investigating the similarity in node locations between the two carriers. This analysis helps to understand how geographical proximity contribute to potential collaboration opportunities.

Initially, the entire spatial area in the dataset is divided into grids of a predetermined size, referred to as the ”grid size.” Each grid represents a distinct two-dimensional spatial region. This process facilitates the spatial partitioning necessary to analyze the distribution of nodes among carriers.

For each carrier, the number of nodes within each grid is counted, resulting in a distribution matrix for each carrier that reflects spatial distribution of nodes within the defined grids. This step transforms spatial data into a quantifiable format suitable for further analysis.

After that, to explore the similarity between the two carriers, cosine similarity is employed due to its simplicity in comparing distribution similarities and its efficiency in short-time computations (Xia et al., 2015). The formula for cosine similarity is as follows. x and x' refer to the two vectors.

$$\text{Cosine Similarity} = \frac{\sum_{i=1}^{d} x_i \times x'_i}{\sqrt{\sum_{i=1}^{d} x_i^2} \times \sqrt{\sum_{i=1}^{d} x_i'^2}} \tag{6.2}$$

This formula measures the directional similarity between two vectors, regardless of their magnitude. By focusing on direction rather than length, cosine similarity emphasizes proportional relationships between node distributions. Before computation, each distribution matrix is flattened into a one-dimensional vector. The relationship between cosine similarity and the collaboration index is illustrated in Figure 6.6.

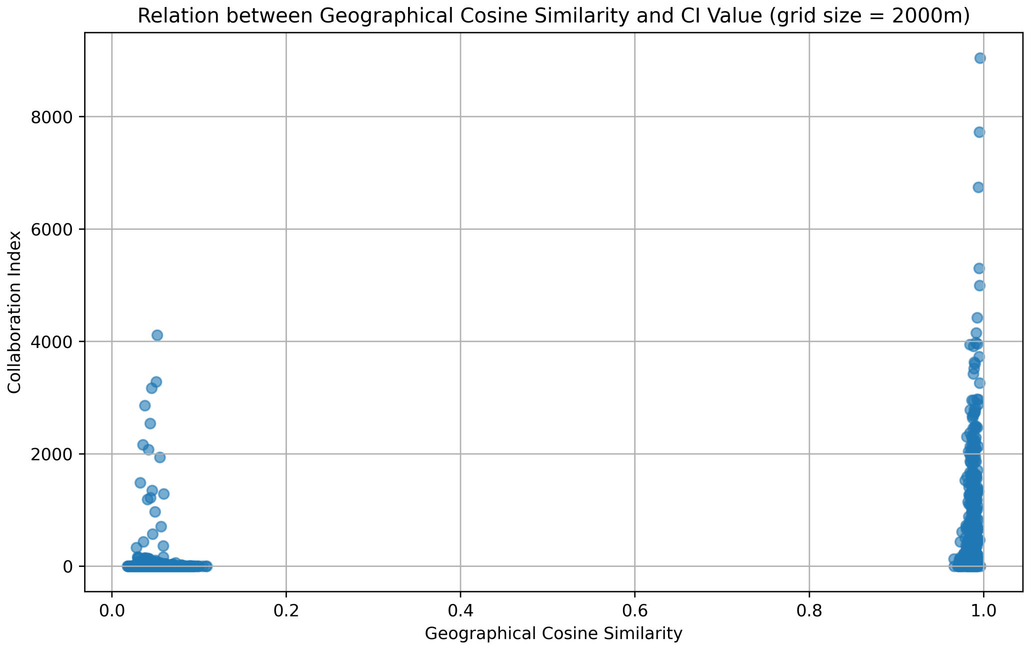


Figure 6.6: Relationship between geographical cosine similarity and collaboration index across carrier pairs (grid size = 2000m)

The cosine similarity values are computed for all carrier pairs to investigate the correlation with their collaboration index. The Pearson correlation is applied to measure the linear relationship between cosine similarity and the collaboration index (CI), indicating the relationship between the geographical location of nodes and the collaboration index.

Several different grid size are given, and all the correlation coefficient are shown in Table 6.3. For the grid size 500, 1000, 2000 and 3000, the result of coefficient are about 0.50, indicating a moderate positive relationship between cosine similarity and the CI value. A positive correlation coefficient suggests that as cosine similarity increases, the CI value tends to increase as well. This implies that carriers with similar spatial distributions may also have higher collaboration potential. All of the coefficient falls within the range of 0.3 to 0.7, which is typically considered a moderate positive correlation. While this is not a very strong correlation, it does indicate a noticeable connection between the two variables.

Table 6.3: Correlation coefficients between geographical cosine similarity and collaboration index at various grid sizes

Grid size (m)	500	1000	2000	3000	5000
Correlation coefficient	0.57	0.54	0.51	0.49	0.39

Vehicle type relation

Examining vehicle types allows for an assessment of how vehicle characteristics influence the collaboration index. This analysis explores whether the types of vehicles owned by carriers in the suggesting collaboration plan are related to the collaboration index.

Firstly, the same vehicle rate between carriers is computed to show the percentage of same vehicles in the collaboration plans between specific carrier pairs. A higher same vehicle rate indicates a greater number of collaborations suggested between trips operated by the same types of vehicles. Figure 6.7 shows the result. The Pearson correlation coefficient is also applied here to examine the relationship between the same vehicle rate and the collaboration index. The result of the Pearson correlation coefficient is -0.028, suggesting a very weak negative linear relationship between the two variables. The value being close to zero implies that there is almost no linear correlation, which means the same vehicle rate are not linearly related with the CI.

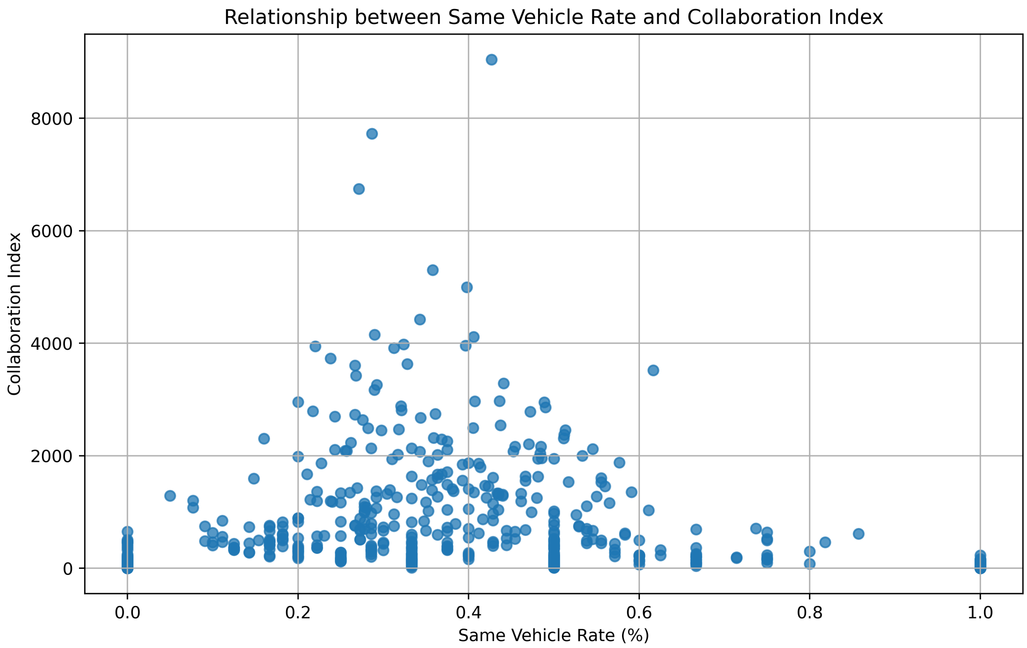


Figure 6.7: Relationship between same types of vehicle rate and collaboration index across carrier pairs

7 Discussion

This research focused on the area of partner selection in road freight transport horizontal collaboration, addressing a notable gap in the current literature. Previous work has often neglected the operational domain, specifically the selection of partners based on freight transportation data. While many studies employ complex optimization approaches to determine collaboration plans between two participants, there has been less emphasis on identifying potential collaborations from a broader perspective.

7.1 Research summary

In addressing the research questions outlined in Section 1.4, each question is resolved as follows.

RQ1: How can the 'collaboration index' be used by companies and authorities for partner selection and evaluation in freight collaboration transportation?

For sub-question 1, the 'collaboration index' offers significant advantages to various stakeholders. For carriers participating in collaboration, the index facilitates the efficient selection of partners with high expected benefits. The Collaborative Decision Support System further provides detailed collaborative plans for each carrier, including vehicle routes, modes of collaboration, and loading or unloading operations for vehicles. This ensures that carriers can optimize their resources and achieve cost savings through strategic collaborations.

For governments and authorities, the collaboration index serves as a valuable reference for assessing the extent of collaboration within the freight transportation sector. It provides insights into current collaboration levels and identifies areas with potential for increased cooperative efforts, thereby informing policy and regulatory decisions to support and enhance freight transport efficiency.

Coordinators, who act as neutral third parties in the Collaborative Transport Networks, can use the collaboration index to manage and promote collaboration opportunities among carriers. By sharing these opportunities based on the collaboration index, coordinators help participants identify the most beneficial partnerships. Additionally, the coordinators ensure the confidentiality of sensitive company information, maintaining trust and neutrality in the collaboration process.

RQ2: What attributes and factors should be included in the 'collaboration index' to reflect the potential for collaborative logistics accurately?

Sub-question 2 would answer from Section 3.1 and Section 3.2. The collaboration determination and the collaboration index formulation involve consideration of various attributes and factors. In the collaboration determination part, several attributes are considered: the distance between trips, the types of cargo, depot locations, vehicle loads, loading and unloading operations, and vehicle capacities. The formulation of the collaboration index incorporates additional variables to enhance the precision and applicability of the index. These include the total profit derived from collaboration, the frequency of collaboration, bias in departure times, limit ranges, the rate of random selection, and the thresholds for cost reduction. All these attributes and factors be considered in the collaboration index research to reflect the potential for collaborative logistics.

RQ3: How can the 'collaboration index' be formulated to quantitatively measure the potential for collaboration?

To address sub-question 3, which seeks to quantitatively identify potential collaboration opportunities among carriers, the collaboration index was formulated. This index was developed through processes of collaboration determination and collaboration index formulation. The collaboration determination process utilized a decision tree to identify four modes of collaboration at the trip level. The collaboration index formulation aggregated these trip-level collaboration modes into a carrier-level quantitative evaluation. Among them, the heuristic local optimization approach was applied to solve the two types of conflicts behind the options for each carrier. As a result, the local optimized collaboration solutions were obtained and the collaboration index quantitatively demonstrated the potential benefits of these feasible collaborative solutions.

7.1.1 Research contribution

The key contribution of this research proposes an innovative collaboration index for carriers, identifying horizontal collaboration opportunities in road freight transport. The collaboration decision tree is employed to determine four modes of collaboration at the trip level from freight transport data. Subsequently, a heuristic local optimization approach aggregates these modes at the carrier level, formulating a quantitative collaboration index.

The findings from the results include sensitivity analysis and further in-depth analysis of the collaboration index. In Section 6.1 three variables were examined for the collaboration index: limit range, random selection rate, and cost reduction threshold. Sensitivity analysis was conducted to assess their effects on the total cost-saving rate. The results indicated that an increase in the limit range could generate more opportunities for collaboration, with a corresponding upward trend in total cost savings. A decrease in the random selection rate resulted in a reduction in the total cost-saving rate. Conversely, as the cost reduction threshold increased, the total cost-saving rate generally declined, with a significant drop observed when the threshold exceeded

30%. This analysis enhances the understanding of the collaboration index by applying it across various scenarios with the three variables.

In this dataset, companies participating in horizontal collaboration can achieve an overall average cost-saving rate of 14.4%. The lowest 25% of companies experience a cost-saving rate of 1.11%, while the highest 25% achieve a cost-saving rate of 22.12%. Further research indicates a moderately positive correlation between the geographical similarity of node locations and the collaboration index among carrier pairs. However, the use of the same vehicle types between carrier pairs does not exhibit a linear relationship with the collaboration index.

7.2 Limitation

There are still some shortages and limitations in this research. Due to the lack of detailed customer information, specifically the trade level production-consumption data, this study only considered weight variations between node visits and did not account for changes in cargo quantity or volume. The research assumed that the cargo carried in a vehicle during its entire tour is homogeneous. Also, although the compatible cargo types were assumed in this research, the application of the compatibility matrix still needs to be examined and investigated through real-world operations.

Additionally, the development of the collaboration index was based on a heuristic local optimization approach, guided by predefined selection metrics. While this method proves efficient, it falls short of achieving global optimization. It does not minimize the total costs incurred by all participants. As a result, the final collaborative solutions for each carrier do not represent globally optimized outcomes. Moreover, the distribution of profits among participants has not been addressed in this research.

7.3 Recommendation

Future studies should aim to utilize operational data alongside customer data, including detailed cargo flows, and account for the volume and quantity of cargo to identify horizontal collaboration opportunities. This comprehensive approach would yield more practical and applicable results for modes of collaboration.

The four collaboration modes identified through the decision tree introduced in Section 3.1.3 not only categorize the collaboration types but also assign labels on trip-level. These labels can be used as target values for machine learning approaches, allowing for further development with additional preferences.

Additionally, this research employed a relatively simple decision-making process to optimize collaborative solutions. Future research could compare this straightforward approach with other optimization models or data-driven methods to evaluate their effectiveness and efficiency. Such comparisons could provide valuable insights into the advantages and limitations of various methodologies in optimizing collaborative freight transport.

The collaboration index introduced in this study also presents an interesting area for further exploration. Future work could investigate its real-world applications, for example, the profit distribution based on the contribution, examining how the collaboration index can be integrated with existing transportation networks. Exploring these could significantly enhance the practical and theoretical understanding of the collaboration index in freight transportation.

8 Conclusion

This research aimed to identify potential horizontal collaboration opportunities with the defined collaboration index. By focusing on partner selection through the analysis of freight transportation data, this study fills a significant gap in the existing literature. The collaboration index was developed to quantitatively assess potential collaboration opportunities among carriers. It utilizes a collaboration decision tree and heuristic local optimization to aggregate data from the trip-level to the carrier-level evaluation. Several attributes were considered, including trip route, depot position, load weight, operational tasks, vehicle types, distance, and cost. The integration of these attributes facilitated the construction of the comprehensive approach. Different scenarios defined by three variables limit range, random selection rate, and cost reduction threshold, influence the cost savings achieved through collaboration. The sensitivity analysis highlights the impact of these variables on cost savings.

According to the research data, if the carriers engage in horizontal collaboration, they have the potential to achieve an average cost-saving of 14.4%, with the top 25% of carriers realizing savings of up to 22.12%. The similarity in geographical node locations between carriers is associated with the value of the collaboration index. However, the same types of vehicles do not appear to have a significant relationship with the collaboration index.

The implications of this research are multifaceted. For carriers, the collaboration index serves as a strategic tool to select partners, potentially leading to optimized operations and cost reductions. The collaboration index has become a fundamental quantitative value that illustrates the potential benefits and cost reductions, serving as a motivation for participants to engage in collaboration. Governments and authorities can leverage the index as a benchmark for the extent of collaboration within road freight transportation, which could guide policy decisions aimed at enhancing transport efficiency. Furthermore, coordinators within Collaborative Transport Networks can utilize the index to suggest and manage beneficial partnerships among carriers, ensuring the neutrality and confidentiality required for successful collaboration.

The research acknowledges its limitations, such as the lack of detailed trade-level data and the reliance on a heuristic local optimization approach that does not achieve global cost minimization. Future research should, therefore, expand to include production-consumption data to investigate the collaboration index further. Comparing the current model with other optimization methods could also provide deeper insights into the collaborative strategies in freight transport.

While the collaboration index developed in this research serves as a quantitative tool for identifying horizontal collaboration in road freight transport, the collaboration index leads to a practical impact on the quantitative identification of collaboration partners, seeking to find partners to receive win-win. The adoption of the collaboration index in real-world applications such as collaboration decision support system (CDSS) has strong potential and the research on collaboration index could be further improved.

The collaboration index developed in this research provides a quantitative framework for identifying potential partners in horizontal collaboration within road freight transport. This index facilitates the strategic pairing of companies aiming to achieve mutually beneficial. When integrated into a Collaboration Decision Support System (CDSS), the collaboration index demonstrates significant potential to enhance real-world applications. Further development and research into the collaboration index would be essential to maximize its efficiency and broaden its applicability in road freight logistics. Such advancements would be expected to pave the way for more efficient road freight transport with horizontal collaboration.

References

Agarwal, R., & Ergun, Ö. (2010). Network Design and Allocation Mechanisms for Carrier Alliances in Liner Shipping [Publisher: INFORMS]. *Operations Research*, *58*(6), 1726–1742. https://doi.org/10.1287/opre.1100.0848

Amer, L. E., & Eltawil, A. B. (2015). Analysis of quantitative models of horizontal collaboration in supply chain network design: Towards "green collaborative" strategies. *2015 International Conference on Industrial Engineering and Operations Management (IEOM)*, 1–10. https://doi.org/10.1109/IEOM.2015.7093759

Audy, J.-F., D'Amours, S., & Rönnqvist, M. (2007). Business Models for Collaborative Planning in Transportation: An Application to Wood Products. In L. M. Camarinha-Matos, H. Afsarmanesh, P. Novais, & C. Analide (Eds.), *Establishing the Foundation of Collaborative Networks* (pp. 667–676). Springer US. https://doi.org/10.1007/978-0-387-73798-0_72

Audy, J.-F., Lehoux, N., D'Amours, S., & Rönnqvist, M. (2012). A framework for an efficient implementation of logistics collaborations. *International Transactions in Operational Research*, *19*(5), 633–657. https://doi.org/10.1111/j.1475-3995.2010.00799.x

Awasthi, A., Adetiloye, T., & Crainic, T. G. (2016). Collaboration partner selection for city logistics planning under municipal freight regulations. *Applied Mathematical Modelling*, *40*(1), 510–525. https://doi.org/10.1016/j.apm.2015.04.058

Bahinipati, B. K., Kanda, A., & Deshmukh, S. G. (2009). Horizontal collaboration in semiconductor manufacturing industry supply chain: An evaluation of collaboration intensity index. *Computers & Industrial Engineering*, *57*(3), 880–895. https://doi.org/10.1016/j.cie.2009.03.003

Beliën, J., Boute, R., Creemers, S., Bruecker, P., Gijsbrechts, J., Padilla Tinoco, S. V., & Verheyen, W. (2017). Collaborative Shipping: Logistics in the Sharing Economy.

Cruijssen, F., Dullaert, W., & Fleuren, H. (2007). Horizontal Cooperation in Transport and Logistics: A Literature Review. *Transportation Journal*, *46*. https://doi.org/10.5325/transportationj.46.3.0022

Dahl, S., & Derigs, U. (2011). Cooperative planning in express carrier networks — An empirical study on the effectiveness of a real-time Decision Support System. *Decision Support Systems*, *51*(3), 620–626. https://doi.org/10.1016/j.dss.2011.02.018

de Bok, M., & Tavasszy, L. (2018). An empirical agent-based simulation system for urban goods transport (MASS-GT). *Procedia Computer Science*, *130*, 126–133. https://doi.org/10.1016/j.procs.2018.04.021

Deng, S., Zhou, D., Wu, G., Wang, L., & You, G. (2023). Evolutionary game analysis of three parties in logistics platforms and freight transportation companies' behavioral strategies for horizontal collaboration considering vehicle capacity utilization. *Complex & Intelligent Systems*, *9*(2), 1617–1637. https://doi.org/10.1007/s40747-022-00873-9

Dennis-Bauer, S., & Jaller, M. (2023). Truck transportation in California: Disaggregating public health costs from criteria pollutants. *Transportation Research Part D: Transport and Environment*, *122*, 103850. https://doi.org/10.1016/j.trd.2023.103850

Dong, X., Wang, Y., Li, X., Zhong, Z., Shen, X., Sun, H., & Hu, B. (2023). Understanding the influencing factors of taxi ride-sharing: A case study of Chengdu, China. *Transportation Research Part A: Policy and Practice*, *176*, 103819. https://doi.org/10.1016/j.tra.2023.103819

Du, H., Wang, Y., Shi, L., & Xing, Y. (2022). Investigating Evaluation Indicators of Intelligent Vehicle Sharing Based on Operation Efficiency: A Case Study in Xiong'an New Area, China. *Journal of Advanced Transportation*, *2022*. https://doi.org/10.1155/2022/7418127

Ergun, O., Kuyzu, G., & Savelsbergh, M. (2007). Reducing Truckload Transportation Costs Through Collaboration [Publisher: INFORMS]. *Transportation Science*, *41*(2), 206–221. https://doi.org/10.1287/trsc.1060.0169

Essam, L., & Eltawil, A. (2014). Collaborative Sustainable Supply Chain Network Design: State Of The Art and Solution Framework.

Eurostat. (2023, October). Road freight transport by journey characteristics. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Road_freight_transport_by_journey_characteristics#Empty_runnings_of_road_freight_vehicles

Feng, B., Fan, Z.-P., & Ma, J. (2010). A method for partner selection of codevelopment alliances using individual and collaborative utilities. *International Journal of Production Economics*, *124*(1), 159–170. https://doi.org/10.1016/j.ijpe.2009.10.019

Ferrell, W., Ellis, K., Kaminsky, P., & Rainwater, C. (2020). Horizontal collaboration: Opportunities for improved logistics planning. *International Journal of Production Research*, *58*(14), 4267–4284. https://doi.org/10.1080/00207543.2019.1651457

Gansterer, M., & Hartl, R. F. (2018). Collaborative vehicle routing: A survey. *European Journal of Operational Research*, *268*(1), 1–12. https://doi.org/10.1016/j.ejor.2017.10.023

Gansterer, M., & Hartl, R. F. (2020). Shared resources in collaborative vehicle routing. *TOP*, *28*(1), 1–20. https://doi.org/10.1007/s11750-020-00541-6

Ghomi, V., Gligor, D., Shokoohyar, S., Alikhani, R., & Nezami, F. G. (2023). An optimization model for collaborative logistics among carriers in vehicle routing problems with cross-docking [Num Pages: 36 Place: Ponte Vedra Beach, United Kingdom Publisher: Emerald Group Publishing Limited]. *International Journal of Logistics Management*, *34*(6), 1700–1735. https://doi.org/10.1108/IJLM-11-2021-0515

Guajardo, M., & Rönnqvist, M. (2016). A review on cost allocation methods in collaborative transportation. *International Transactions in Operational Research*, *23*(3), 371–392. https://doi.org/10.1111/itor.12205

Hernández, S., & Peeta, S. (2014). A carrier collaboration problem for less-than-truckload carriers: Characteristics and carrier collaboration model. *Transportmetrica A: Transport Science*, *10*(4), 327–349. https://doi.org/10.1080/23249935.2013.766279

Hernández, S., Unnikrishnan, A., & Awale, S. S. (2012). Centralized Carrier Collaboration Multihub Location Problem for Less-Than-Truckload Industry: Hybrid Hub-and-Spoke Network. *Transportation Research Record: Journal of the Transportation Research Board*, *2269*(1), 20–28. https://doi.org/10.3141/2269-03

Karam, A., Reinau, K. H., & Østergaard, C. R. (2021). Horizontal collaboration in the freight transport sector: Barrier and decision-making frameworks. *European Transport Research Review*, *13*(1), 53. https://doi.org/10.1186/s12544-021-00512-3

Lai, M., Cai, X., & Hall, N. G. (2022). Cost Allocation for Less-Than-Truckload Collaboration via Shipper Consortium [Publisher: INFORMS]. *Transportation Science*, *56*(3), 585–611. https://doi.org/10.1287/trsc.2021.1066

Luan, J., Daina, N., Reinau, K. H., Sivakumar, A., & Polak, J. W. (2022). A data-based opportunity identification engine for collaborative freight logistics based on a trailer capacity graph. *Expert Systems with Applications*, *210*, 118494. https://doi.org/10.1016/j.eswa.2022.118494

McKinnon, A. C., & Ge, Y. (2006). The potential for reducing empty running by trucks: A retrospective analysis [Publisher: Emerald Group Publishing Limited]. *International Journal of Physical Distribution & Logistics Management*, *36*(5), 391–410. https://doi.org/10.1108/09600030610676268

Mohammed, R. A., Nadi, A., Tavasszy, L., & Bok, M. d. (2023). Data Fusion Approach to Identify Distribution Chain Segments in Freight Shipment Databases [Publisher: SAGE Publications Inc]. *Transportation Research Record*, *2677*(6), 310–323. https://doi.org/10.1177/03611981221147049

Naesens, K., Gelders, L., & Pintelon, L. (2009). A swift response framework for measuring the strategic fit for a horizontal collaborative initiative. *International Journal of Production Economics*, *121*(2), 550–561. https://doi.org/10.1016/j.ijpe.2007.04.004

Pan, S., Trentesaux, D., Ballot, E., & Huang, G. Q. (2019). Horizontal collaborative transport: Survey of solutions and practical implementation issues [Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00207543.2019.1574040]. *International Journal of Production Research*, *57*(15-16), 5340–5361. https://doi.org/10.1080/00207543.2019.1574040

Statista. (2024, January). Transportation emissions in the European Union - Statistics & Facts. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Road_freight_transport_by_journey_characteristics#Empty_runnings_of_road_freight_vehicles

Vargas, A., Patel, S., & Patel, D. (2018). Towards a Business Model Framework to Increase Collaboration in the Freight Industry [Number: 4 Publisher: Multidisciplinary Digital Publishing Institute]. *Logistics*, *2*(4), 22. https://doi.org/10.3390/logistics2040022

Wang, J., Yu, Y., & Tang, J. (2018). Compensation and profit distribution for cooperative green pickup and delivery problem. *Transportation Research Part B: Methodological*, *113*, 54–69. https://doi.org/10.1016/j.trb.2018.05.003

Wang, Y., Li, Q., Guan, X., Xu, M., Liu, Y., & Wang, H. (2021). Two-echelon collaborative multi-depot multi-period vehicle routing problem. *Expert Systems with Applications*, *167*, 114201. https://doi.org/10.1016/j.eswa.2020.114201

Xia, P., Zhang, L., & Li, F. (2015). Learning similarity with cosine similarity ensemble. *Information Sciences*, *307*, 39–52. https://doi.org/10.1016/j.ins.2015.02.024

Appendix A

1 Code for data prepossessing

```python
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
# %%
# Load the datasets
preprocess_file_path = "E:\\1MASTER in TUD\\@Master Thesis\\data\\preprocess_file_filter.csv"
carrying_capacity_path = "E:\\1MASTER in TUD\\@Master Thesis\\Data Description\\CarryingCapacity.csv"

preprocess_df = pd.read_csv(preprocess_file_path)
carrying_capacity_df = pd.read_csv(carrying_capacity_path)

# Display the first few rows of each dataset to understand their structure
preprocess_df.head(), carrying_capacity_df.head()


# Map the carrying capacity to the preprocess dataset based on VEHTYPE and ID
capacity_mapping = carrying_capacity_df.set_index('ID')['ton'].to_dict()
preprocess_df['Capacity'] = preprocess_df['VEHTYPE'].map(capacity_mapping)

# Calculate the Loadrate as TRIP_WEIGHT divided by Capacity
preprocess_df['Loadrate'] = preprocess_df['TRIP_WEIGHT'] / preprocess_df['Capacity']

# Show the updated dataframe to verify the changes
preprocess_df[['VEHTYPE', 'Capacity', 'TRIP_WEIGHT', 'Loadrate']].head()
#%%
preprocess_df.rename(columns={'Pick (demand origin)': 'Pick_demand_origin'}, inplace=True)
#%%
#
preprocess_df['NODE_COUNT'] = preprocess_df['TRIP_ID'].str.split('_').str[-1].astype(int) + 1

preprocess_df['NODE_COUNT'] = preprocess_df.groupby('TOUR_ID')['NODE_COUNT'].transform('max')
#%%

```

```
preprocess_df['Distance'] = np.sqrt((preprocess_df['X_DEST'] -
    preprocess_df['X_ORIG'])**2 + (preprocess_df['Y_DEST'] -
    preprocess_df['Y_ORIG'])**2)

#%%
# Save the updated dataframe to a new CSV file
updated_file_path = "E:\\1MASTER in TUD\\@Master Thesis\\data\\
    preprocess_file_filter_capacity.csv"
preprocess_df.to_csv(updated_file_path, index=False)
```

2 Code for collaboration determination

```
import pandas as pd
from itertools import combinations
import numpy as np
import math
from scipy.sparse import lil_matrix


import time

# Record the start time
start_time = time.time()
import pandas as pd
import geopandas as gpd
from shapely.geometry import Point

# Load data
df = pd.read_csv("E:\\1MASTER in TUD\\@Master Thesis\\data\\
    preprocess_file_filter_capacity.csv")

def create_geodataframe(df):
    # Convert latitude and longitude to point geometry objects
    # Load and prepare data
    df['X'] = df.apply(lambda row: (row['X_ORIG'], row['X_DEST']), axis
    =1)
    df['Y'] = df.apply(lambda row: (row['Y_ORIG'], row['Y_DEST']), axis
    =1)

    gdf = gpd.GeoDataFrame(df, geometry=[Point(xy) for xy in zip(df.X,
    df.Y)])
    gdf.set_crs(epsg=28992, inplace=True)  # Ensure the correct
    coordinate system is used
    return gdf

def find_close_pairs(gdf, max_distance=0.1): # Change the Limit of range
    # Create spatial index
    spatial_index = gdf.sindex
    pairs_list = []
    seen_pairs = set()  # Used to store already processed TRIP_ID pairs
    for idx, row in gdf.iterrows():
        # Use buffer and spatial index to find nearby points
```

```
        possible_matches_index = list(spatial_index.intersection(row.geometry.buffer(max_distance).bounds))
        possible_matches = gdf.iloc[possible_matches_index]
        precise_matches = possible_matches[possible_matches.geometry.distance(row.geometry) <= max_distance]

        # Add qualifying trip pairs
        for idx2, match_row in precise_matches.iterrows():
            if idx != idx2 and row['CARRIER_ID'] != match_row['CARRIER_ID']:  # Avoid pairing a trip with itself and with different CARRIER_ID
                pair = tuple(sorted([row['TRIP_ID'], match_row['TRIP_ID']]))  # Create and sort TRIP_ID pairs
                if pair not in seen_pairs and (row['ORIG'] != row['DEST'] and match_row['ORIG'] != match_row['DEST']):
                    seen_pairs.add(pair)
                    # Create point geometry objects for origin and destination
                    point_a_orig = Point(row['X_ORIG'], row['Y_ORIG'])
                    point_b_orig = Point(match_row['X_ORIG'], match_row['Y_ORIG'])
                    point_a_dest = Point(row['X_DEST'], row['Y_DEST'])
                    point_b_dest = Point(match_row['X_DEST'], match_row['Y_DEST'])

                    # Calculate the distance between origins and destinations
                    orig_distance = point_a_orig.distance(point_b_orig)
                    dest_distance = point_a_dest.distance(point_b_dest)
                    if orig_distance <= max_distance and dest_distance <= max_distance:
                        pairs_list.append({
                            'PAIR_ID': f"({row['CARRIER_ID']}, {match_row['CARRIER_ID']})",
                            'TRIP_ID_A': row['TRIP_ID'],
                            'TRIP_ID_B': match_row['TRIP_ID'],
                            'ORIG_A': row['ORIG'],
                            'ORIG_B': match_row['ORIG'],
                            'DEST_A': row['DEST'],
                            'DEST_B': match_row['DEST'],
                            'ORIG_DISTANCE': orig_distance,
                            'DEST_DISTANCE': dest_distance,
                            'DISTANCE_A': row['Distance'],
                            'DISTANCE_B': match_row['Distance'],
                            'DC_ID_A': row['DC_ID'],
                            'DC_ID_B': match_row['DC_ID'],
                            'NODE_COUNT_A': row['NODE_COUNT'],
                            'NODE_COUNT_B': match_row['NODE_COUNT'],
                            'VEHTYPE_A': row['VEHTYPE'],
                            'VEHTYPE_B': match_row['VEHTYPE'],
                            'N_SHIP_A': row['N_SHIP'],
                            'N_SHIP_B': match_row['N_SHIP'],
                            'TRIP_DEPTIME_A': row['TRIP_DEPTIME'],
                            'TRIP_DEPTIME_B': match_row['TRIP_DEPTIME'],
                            'TRIP_ARRTIME_A': row['TRIP_ARRTIME'],
```

```
                                'TRIP_ARRTIME_B': match_row['TRIP_ARRTIME'],
                                'TRIP_WEIGHT_A': row['TRIP_WEIGHT'],
                                'TRIP_WEIGHT_B': match_row['TRIP_WEIGHT'],
                                'LOG_SEG': f"({row['LOG_SEG']}, {match_row['LOG_SEG']})",  # Use compa_matrix.iloc[0,2] to find
                                'O_Pick_A': row['Pick_demand_origin'],
                                'O_Pick_B': match_row['Pick_demand_origin'],
                                'Capacity_A': row['Capacity'],
                                'Capacity_B': match_row['Capacity'],
                                'Loadrate_A': row['Loadrate'],
                                'Loadrate_B': match_row['Loadrate']

                            })

    return pd.DataFrame(pairs_list)

# Load and prepare data
gdf = create_geodataframe(df)

# Find trip pairs with a distance less than % meters
filtered_pairs = find_close_pairs(gdf)

# Record the end time
end_time = time.time()

# Calculate runtime
elapsed_time = end_time - start_time
print(f"Runtime: {elapsed_time} seconds")

# Convert results to DataFrame
pairs_df = pd.DataFrame(filtered_pairs)

# Load compatibility matrix for logistics segment
compa_matrix = pd.read_csv("E:\\1MASTER in TUD\\@Master Thesis\\data\\Compa_cargo.csv"
                            ,header=None,index_col=False, usecols=range(9))
# Compatible logistics pair
compa_pairs_df = pairs_df.iloc[[i for i, log_seg in enumerate(pairs_df['LOG_SEG'])
                                   if compa_matrix.iloc[eval(log_seg)] != 0]]

updated_file_path = "E:\\1MASTER in TUD\\@Master Thesis\\data\\range_considered\\range_0_trip_pairs.csv"
compa_pairs_df.to_csv(updated_file_path, index=False)

#%%
# Load the file
compa_pairs_df = pd.read_csv("E:\\1MASTER in TUD\\@Master Thesis\\data\\range_considered\\range_0_trip_pairs.csv")

# Initialize empty lists to store the labels for each row
collaboration_status = []
depot_status = []
```

```python

for index, row in compa_pairs_df.iterrows():
    # Define a label for whether to execute my logic
    execute_logic = False

    O_plus_A = row['O_Pick_A'] >= 0
    O_plus_B = row['O_Pick_B'] >= 0

    O_matches_A = row['ORIG_A'] == row['DC_ID_A']
    O_matches_B = row['ORIG_B'] == row['DC_ID_B']

    approved_A_carry_B = False
    approved_B_carry_A = False
    both_approved = False
    O_depot_A = False
    O_depot_B = False
    none_approved = False

    if O_matches_A and O_matches_B:
        if (row['NODE_COUNT_A'] > 2):
            execute_logic = False
        elif (row['NODE_COUNT_B'] > 2):
            execute_logic = False
        else:
            execute_logic = True
            O_depot_A = True
            O_depot_B = True
    elif (O_matches_A != O_matches_B):
        if (O_matches_A):
            if (row['NODE_COUNT_A'] > 2):
                execute_logic = False
            else:
                O_depot_A = True
                execute_logic = True
        elif (O_matches_B):
            if (row['NODE_COUNT_B'] > 2):
                execute_logic = False
            else:
                O_depot_B = True
                execute_logic = True
        else:
            execute_logic = False
    else:
        execute_logic = True


    if execute_logic:
        if O_plus_A and O_plus_B:
            #A carries cargo from B, B skips node and saves distance.
            condition_A = (row['TRIP_WEIGHT_A'] + row['O_Pick_B']) <=
   row['Capacity_A']
            #B carries cargo from A, A skips node and saves distance.
            condition_B = (row['TRIP_WEIGHT_B'] + row['O_Pick_A']) <=
   row['Capacity_B']

```

```
            if condition_A and condition_B:
                # Both available need average to show benefits
                both_approved = True
            elif condition_A:
                approved_A_carry_B = True
            elif condition_B:
                approved_B_carry_A = True
            else:
                none_approved = True

        elif not O_plus_A and not O_plus_B:
            none_approved = True


        elif (O_plus_A != O_plus_B):
            if O_plus_A:
                # B carries A
                if ((row['TRIP_WEIGHT_B'] + row['O_Pick_A']) <= row['Capacity_B']):
                    approved_B_carry_A = True
                else:
                    none_approved = True
            elif O_plus_B:
                if ((row['TRIP_WEIGHT_A'] + row['O_Pick_B']) <= row['Capacity_A']):
                    approved_A_carry_B = True
                else:
                    none_approved = True
            else:
                none_approved = True

        if not (approved_A_carry_B or approved_B_carry_A or both_approved):
            none_approved = True

    else:
        none_approved = True

    # Set the final label for this row based on the label variables
    if both_approved:
        collaboration_status.append('both_approved')
    elif approved_A_carry_B:
        collaboration_status.append('approved_A_carry_B')
    elif approved_B_carry_A:
        collaboration_status.append('approved_B_carry_A')
    elif none_approved:
        collaboration_status.append('none_approved')
    else:
        collaboration_status.append('undetermined')  # As a safety measure in case no label is set

    if (O_depot_A != O_depot_B):
        if O_depot_A:
            depot_status.append('Origin_depot_A')
        elif O_depot_B:
```

```
                depot_status.append('Origin_depot_B')
            else:
                depot_status.append('not-same-none')
        elif O_depot_A and O_depot_B:
            depot_status.append('Origin_depot_both_AB')
        else:
            depot_status.append('none')

# Add the collaboration status list to the DataFrame as a new column
compa_pairs_df['Collaboration_Status'] = collaboration_status
compa_pairs_df['Depot_Status'] = depot_status

# Check if there are any rows with 'Collaboration_Status' equal to '
    undetermined'
has_undetermined = compa_pairs_df['Collaboration_Status'].eq('
    undetermined').any()
# Print the result
print(f"Are there any 'undetermined': {has_undetermined}")
# Save the file
updated_file_path = "E:\\1MASTER in TUD\\@Master Thesis\\data\\
    range_considered\\range_0_Colla_status.csv"
compa_pairs_df.to_csv(updated_file_path, index=False)
```

3 Code for cost and benefit

```
import pandas as pd
from itertools import combinations
import numpy as np
import math
from scipy.sparse import lil_matrix

# Load the uploaded CSV file
df = pd.read_csv("E:\\1MASTER in TUD\\@Master Thesis\\data\\
    preprocess_file_filter_capacity.csv")

# Load the file
Filter_compa_pairs_df = pd.read_csv("E:\\1MASTER in TUD\\@Master Thesis
    \\data\\range_considered\\range_0_Colla_status.csv")


# Generate new ID pointing to the previous TRIP_ID
def generate_prev_id(trip_id):
    parts = trip_id.split('_')
    last_part = int(parts[-1])  # Get the last number
    if last_part == 0:
        return None
    new_last_part = str(last_part - 1)  # Generate the new last number,
    pointing to the previous row
    new_trip_id = '_'.join(parts[:-1] + [new_last_part])  # Generate the
     new TRIP_ID
    return new_trip_id

# Apply the function to the TRIP_ID_A and TRIP_ID_B columns
```

```
Filter_compa_pairs_df['Prev_TRIP_ID_A'] = Filter_compa_pairs_df['
    TRIP_ID_A'].apply(generate_prev_id)
Filter_compa_pairs_df['Prev_TRIP_ID_B'] = Filter_compa_pairs_df['
    TRIP_ID_B'].apply(generate_prev_id)

# Create a mapping from df's TRIP_ID to ORIG
trip_id_to_orig = pd.Series(df.ORIG.values, index=df.TRIP_ID).to_dict()
# Use the mapping and Prev_TRIP_ID_A to fill Prev_ORIG_A, same for B
Filter_compa_pairs_df['Prev_ORIG_A'] = Filter_compa_pairs_df['
    Prev_TRIP_ID_A'].map(trip_id_to_orig)
Filter_compa_pairs_df['Prev_ORIG_B'] = Filter_compa_pairs_df['
    Prev_TRIP_ID_B'].map(trip_id_to_orig)

# Create a mapping from df's TRIP_ID to Prev_TRIP_WEIGHT_A
trip_id_to_weight = pd.Series(df.TRIP_WEIGHT.values, index=df.TRIP_ID).
    to_dict()
# Similarly
Filter_compa_pairs_df['Prev_TRIP_WEIGHT_A'] = Filter_compa_pairs_df['
    Prev_TRIP_ID_A'].map(trip_id_to_weight)
Filter_compa_pairs_df['Prev_TRIP_WEIGHT_B'] = Filter_compa_pairs_df['
    Prev_TRIP_ID_B'].map(trip_id_to_weight)

# Create a mapping from df's TRIP_ID to Prev_Dist_A
trip_id_to_Dist = pd.Series(df.Distance.values, index=df.TRIP_ID).
    to_dict()
# Similarly
Filter_compa_pairs_df['Prev_Dist_A'] = Filter_compa_pairs_df['
    Prev_TRIP_ID_A'].map(trip_id_to_Dist)
Filter_compa_pairs_df['Prev_Dist_B'] = Filter_compa_pairs_df['
    Prev_TRIP_ID_B'].map(trip_id_to_Dist)

#approved_A_carry_B B has more than two visit points
Slice_compa_pairs_df = Filter_compa_pairs_df[(Filter_compa_pairs_df['
    Collaboration_Status'] == 'approved_A_carry_B') & (
    Filter_compa_pairs_df['NODE_COUNT_B'] > 1) ]

# Calculate After_Colla_B need first find the new distance from
    Prev_ORIG_B to DEST (skip the ORIG)
# Ensure df uses TRIP_ID as the index
if df.index.name != 'TRIP_ID':
     df.set_index('TRIP_ID', inplace=True)

# Calculate the new distance
# Find the positions of carrier B's two points
# Directly create a mapping dictionary

trip_id_to_coords = {trip_id: tuple(df.loc[trip_id, ['X_DEST', 'Y_DEST'
    ]]) for trip_id in Slice_compa_pairs_df['TRIP_ID_B'] if trip_id in df
    .index}
prev_trip_id_to_coords = {trip_id: tuple(df.loc[trip_id, ['X_ORIG', '
    Y_ORIG']]) for trip_id in Slice_compa_pairs_df['Prev_TRIP_ID_B'] if
    trip_id in df.index}

def euclidean_distance(coord1, coord2):
     """Calculate the Euclidean distance between two coordinates"""
```

```
    x1, y1 = coord1
    x2, y2 = coord2
    return math.sqrt((x2 - x1)**2 + (y2 - y1)**2)

def get_distance(trip_id, prev_trip_id):
    """Get the Euclidean distance between the coordinates corresponding
   to two trip IDs"""
    coord1 = trip_id_to_coords.get(trip_id)
    coord2 = prev_trip_id_to_coords.get(prev_trip_id)
    if coord1 is not None and coord2 is not None:
        return euclidean_distance(coord1, coord2)
    else:
        return None  # If either trip ID does not exist, return None

# Calculate the distances
new_distances = []
for trip_id, prev_trip_id in zip(Slice_compa_pairs_df['TRIP_ID_B'],
   Slice_compa_pairs_df['Prev_TRIP_ID_B']):
    distance = get_distance(trip_id, prev_trip_id)
    new_distances.append(distance)

Filter_compa_pairs_df['NEW_Distance_B'] = np.nan  # Initialize all
   distances as NaN
Slice_compa_pairs_df['NEW_Distance_B'] = new_distances
Filter_compa_pairs_df.loc[Slice_compa_pairs_df.index, 'NEW_Distance_B']
   = new_distances

# Set the condition: 'Collaboration_Status' is 'approved_A_carry_B' and
   'NODE_COUNT_B' is 2
condition = (Filter_compa_pairs_df['Collaboration_Status'] == '
   approved_A_carry_B') & (Filter_compa_pairs_df['NODE_COUNT_B'] == 2)
# Save route, distance is 0
Filter_compa_pairs_df.loc[condition, 'NEW_Distance_B'] = 0


#approved_B_carry_A A has more than two visit points
Slice_compa_pairs_df = Filter_compa_pairs_df[(Filter_compa_pairs_df['
   Collaboration_Status'] == 'approved_B_carry_A') ]

# Calculate After_Colla_A need first find the new distance from
   Prev_ORIG_A to DEST (skip the ORIG)
# Ensure df uses TRIP_ID as the index
if df.index.name != 'TRIP_ID':
    df.set_index('TRIP_ID', inplace=True)

# Calculate the new distance
# Find the positions of carrier A's two points
# Directly create a mapping dictionary

trip_id_to_coords = {trip_id: tuple(df.loc[trip_id, ['X_DEST', 'Y_DEST'
   ]]) for trip_id in Slice_compa_pairs_df['TRIP_ID_A'] if trip_id in df
   .index}
prev_trip_id_to_coords = {trip_id: tuple(df.loc[trip_id, ['X_ORIG', '
   Y_ORIG']]) for trip_id in Slice_compa_pairs_df['Prev_TRIP_ID_A'] if
   trip_id in df.index}
```

```python


# Calculate the distances
new_distances = []
for trip_id, prev_trip_id in zip(Slice_compa_pairs_df['TRIP_ID_A'],
    Slice_compa_pairs_df['Prev_TRIP_ID_A']):
     distance = get_distance(trip_id, prev_trip_id)
     new_distances.append(distance)

Filter_compa_pairs_df['NEW_Distance_A'] = np.nan  # Initialize all
    distances as NaN
Slice_compa_pairs_df['NEW_Distance_A'] = new_distances
# Filter_compa_pairs_df.loc[Slice_compa_pairs_df.index, 'NEW_Distance_A
    '] = new_distances
Filter_compa_pairs_df.loc[Slice_compa_pairs_df.index, 'NEW_Distance_A']
    = Slice_compa_pairs_df['NEW_Distance_A']

# Set the condition: 'Collaboration_Status' is 'approved_B_carry_A' and
    'NODE_COUNT_B' is 2
condition = (Filter_compa_pairs_df['Collaboration_Status'] == '
    approved_B_carry_A') & (Filter_compa_pairs_df['NODE_COUNT_A'] == 2)
# Save route, distance is 0
Filter_compa_pairs_df.loc[condition, 'NEW_Distance_A'] = 0

#both_approved considering approved_A_carry_B B has more than two visit
    points
Slice_compa_pairs_df = Filter_compa_pairs_df[(Filter_compa_pairs_df['
    Collaboration_Status'] == 'both_approved') ]

# Calculate After_Colla_B need first find the new distance from
    Prev_ORIG_B to DEST (skip the ORIG)
# Ensure df uses TRIP_ID as the index
if df.index.name != 'TRIP_ID':
     df.set_index('TRIP_ID', inplace=True)

# Calculate the new distance
# Find the positions of carrier B's two points
# Directly create a mapping dictionary

trip_id_to_coords = {trip_id: tuple(df.loc[trip_id, ['X_DEST', 'Y_DEST'
    ]]) for trip_id in Slice_compa_pairs_df['TRIP_ID_B'] if trip_id in df
    .index}
prev_trip_id_to_coords = {trip_id: tuple(df.loc[trip_id, ['X_ORIG', '
    Y_ORIG']]) for trip_id in Slice_compa_pairs_df['Prev_TRIP_ID_B'] if
    trip_id in df.index}


def euclidean_distance(coord1, coord2):
     """Calculate the Euclidean distance between two coordinates"""
     x1, y1 = coord1
     x2, y2 = coord2
     return math.sqrt((x2 - x1)**2 + (y2 - y1)**2)

def get_distance(trip_id, prev_trip_id):
     """Get the Euclidean distance between the coordinates corresponding
```

```
    to two trip IDs"""
    coord1 = trip_id_to_coords.get(trip_id)
    coord2 = prev_trip_id_to_coords.get(prev_trip_id)
    if coord1 is not None and coord2 is not None:
        return euclidean_distance(coord1, coord2)
    else:
        return None  # If either trip ID does not exist, return None

# Calculate the distances
new_distances = []
for trip_id, prev_trip_id in zip(Slice_compa_pairs_df['TRIP_ID_B'], Slice_compa_pairs_df['Prev_TRIP_ID_B']):
    distance = get_distance(trip_id, prev_trip_id)
    new_distances.append(distance)

# Filter_compa_pairs_df['NEW_Distance_B'] = np.nan  # Initialize all distances as NaN
Slice_compa_pairs_df['NEW_Distance_B'] = new_distances
# Filter_compa_pairs_df.loc[Slice_compa_pairs_df.index, 'NEW_Distance_B'] = new_distances
Filter_compa_pairs_df.loc[Slice_compa_pairs_df.index, 'NEW_Distance_B'] = Slice_compa_pairs_df['NEW_Distance_B']


# Set the condition: 'Collaboration_Status' is 'approved_A_carry_B' and 'NODE_COUNT_B' is 2
condition = (Filter_compa_pairs_df['Collaboration_Status'] == 'both_approved') & (Filter_compa_pairs_df['NODE_COUNT_B'] == 2)

# Save route, distance is 0
Filter_compa_pairs_df.loc[condition, 'NEW_Distance_B'] = 0

#Second case both_approved approved_B_carry_A A has more than two visit points
Slice_compa_pairs_df = Filter_compa_pairs_df[(Filter_compa_pairs_df['Collaboration_Status'] == 'both_approved') ]

# Calculate After_Colla_A need first find the new distance from Prev_ORIG_A to DEST (skip the ORIG)
# Ensure df uses TRIP_ID as the index
if df.index.name != 'TRIP_ID':
    df.set_index('TRIP_ID', inplace=True)
# Calculate the new distance
# Find the positions of carrier A's two points
# Directly create a mapping dictionary

trip_id_to_coords = {trip_id: tuple(df.loc[trip_id, ['X_DEST', 'Y_DEST']]) for trip_id in Slice_compa_pairs_df['TRIP_ID_A'] if trip_id in df.index}
prev_trip_id_to_coords = {trip_id: tuple(df.loc[trip_id, ['X_ORIG', 'Y_ORIG']]) for trip_id in Slice_compa_pairs_df['Prev_TRIP_ID_A'] if trip_id in df.index}


# Calculate the distances
```

```
new_distances = []
for trip_id, prev_trip_id in zip(Slice_compa_pairs_df['TRIP_ID_A'],
    Slice_compa_pairs_df['Prev_TRIP_ID_A']):
    distance = get_distance(trip_id, prev_trip_id)
    new_distances.append(distance)

# Filter_compa_pairs_df['NEW_Distance_A'] = np.nan  # Initialize all
    distances as NaN
Slice_compa_pairs_df['NEW_Distance_A'] = new_distances
# Filter_compa_pairs_df.loc[Slice_compa_pairs_df.index, 'NEW_Distance_A
    '] = new_distances
Filter_compa_pairs_df.loc[Slice_compa_pairs_df.index, 'NEW_Distance_A']
    = Slice_compa_pairs_df['NEW_Distance_A']

# Set the condition: 'Collaboration_Status' is 'approved_B_carry_A' and
    'NODE_COUNT_B' is 2
condition = (Filter_compa_pairs_df['Collaboration_Status'] == '
    both_approved') & (Filter_compa_pairs_df['NODE_COUNT_A'] == 2)
# Save route, distance is 0
Filter_compa_pairs_df.loc[condition, 'NEW_Distance_A'] = 0

#Save the file
updated_file_path = "E:\\1MASTER in TUD\\@Master Thesis\\data\\
    range_considered\\range_0_Ready4cost.csv"
Filter_compa_pairs_df.to_csv(updated_file_path, index=False)

#%% Method for cost

import pandas as pd

# Load the dataframes

Cost_df = pd.read_csv("E:\\1MASTER in TUD\\@Master Thesis\\data\\
    range_considered\\range_0_Ready4cost.csv")
ref = pd.read_csv("E:\\1MASTER in TUD\\@Master Thesis\\data\\
    Cost_VehType_2016.csv")

def KMcost(row, vehicle_type_col, distance_col):
    # Extract vehicle type and distance from the row
    vehicle_type = row[vehicle_type_col]

    distance = row[distance_col] if pd.notna(row[distance_col]) else 0
    # distance = row[distance_col]

    # Find the cost per km for the given vehicle type in the ref
    dataframe
    cost_per_km = ref.at[vehicle_type, 'CostPerKm']
    # Calculate the total cost
    costKM = cost_per_km * distance / 1000
    return costKM

#approved_A_carry_B
sli_df = Cost_df[(Cost_df['Collaboration_Status'] == 'approved_A_carry_B
    ')]
# Calculate Before_Colla
```

```
# Calculate energy consumption, directly sum the results, as NaN is already handled internally
sli_df['Before_Colla_A'] = sli_df.apply(
    lambda row: KMcost(row, 'VEHTYPE_A', 'Prev_Dist_A') +
                KMcost(row, 'VEHTYPE_A', 'DISTANCE_A'),
    axis=1
)
sli_df['Before_Colla_B'] = sli_df.apply(
    lambda row: KMcost(row, 'VEHTYPE_B', 'Prev_Dist_B') +
                KMcost(row, 'VEHTYPE_B', 'DISTANCE_B'),
    axis=1
)
# Calculate After_Colla_A
sli_df['After_Colla_A'] = sli_df.apply(
    lambda row: KMcost(row, 'VEHTYPE_A', 'Prev_Dist_A') +
                KMcost(row, 'VEHTYPE_A', 'ORIG_DISTANCE') +
                KMcost(row, 'VEHTYPE_A', 'DISTANCE_B') +
                KMcost(row, 'VEHTYPE_A', 'DEST_DISTANCE') ,
    axis=1
)
sli_df['After_Colla_B'] = sli_df.apply(
    lambda row: KMcost(row, 'VEHTYPE_B', 'NEW_Distance_B'),
    axis=1
)
# Then, update the calculated results back to Cost_df
Cost_df.loc[sli_df.index, 'KMCost_Before_A'] = sli_df['Before_Colla_A']
Cost_df.loc[sli_df.index, 'KMCost_Before_B'] = sli_df['Before_Colla_B']
# Then, update the calculated results back to Cost_df
Cost_df.loc[sli_df.index, 'KMCost_After_A'] = sli_df['After_Colla_A']
Cost_df.loc[sli_df.index, 'KMCost_After_B'] = sli_df['After_Colla_B']

#approved_B_carry_A
sli_df = Cost_df[(Cost_df['Collaboration_Status'] == 'approved_B_carry_A
    ')]
# Calculate Before_Colla
# Calculate energy consumption, directly sum the results, as NaN is already handled internally
sli_df['Before_Colla_A'] = sli_df.apply(
    lambda row: KMcost(row, 'VEHTYPE_A', 'Prev_Dist_A') +
                KMcost(row, 'VEHTYPE_A', 'DISTANCE_A'),
    axis=1
)
sli_df['Before_Colla_B'] = sli_df.apply(
    lambda row: KMcost(row, 'VEHTYPE_B', 'Prev_Dist_B') +
                KMcost(row, 'VEHTYPE_B', 'DISTANCE_B'),
    axis=1
)

# Calculate After_Colla
sli_df['After_Colla_A'] = sli_df.apply(
    lambda row: KMcost(row, 'VEHTYPE_A', 'NEW_Distance_A'),
    axis=1
)
sli_df['After_Colla_B'] = sli_df.apply(
    lambda row: KMcost(row, 'VEHTYPE_B', 'Prev_Dist_B') +
```

```
                KMcost(row, 'VEHTYPE_B', 'ORIG_DISTANCE') +
                KMcost(row, 'VEHTYPE_B', 'DISTANCE_A') +
                KMcost(row, 'VEHTYPE_B', 'DEST_DISTANCE') ,
    axis=1
)
# Then, update the calculated results back to Cost_df
Cost_df.loc[sli_df.index, 'KMCost_Before_A'] = sli_df['Before_Colla_A']
Cost_df.loc[sli_df.index, 'KMCost_Before_B'] = sli_df['Before_Colla_B']
# Then, update the calculated results back to Cost_df
Cost_df.loc[sli_df.index, 'KMCost_After_A'] = sli_df['After_Colla_A']
Cost_df.loc[sli_df.index, 'KMCost_After_B'] = sli_df['After_Colla_B']

#Both_approved
sli_df = Cost_df[(Cost_df['Collaboration_Status'] == 'both_approved')]
# Calculate Before_Colla
sli_df['Before_Colla_A'] = sli_df.apply(
    lambda row: KMcost(row, 'VEHTYPE_A', 'Prev_Dist_A') +
                KMcost(row, 'VEHTYPE_A', 'DISTANCE_A'),
    axis=1
)
sli_df['Before_Colla_B'] = sli_df.apply(
    lambda row: KMcost(row, 'VEHTYPE_B', 'Prev_Dist_B') +
                KMcost(row, 'VEHTYPE_B', 'DISTANCE_B'),
    axis=1
)

# Calculate After_Colla
# First case
sli_df['A-carry-B-After_Colla_A'] = sli_df.apply(
    lambda row: KMcost(row, 'VEHTYPE_A', 'Prev_Dist_A') +
                KMcost(row, 'VEHTYPE_A', 'ORIG_DISTANCE') +
                KMcost(row, 'VEHTYPE_A', 'DISTANCE_B') +
                KMcost(row, 'VEHTYPE_A', 'DEST_DISTANCE') ,

    axis=1
)
sli_df['A-carry-B-After_Colla_B'] = sli_df.apply(
    lambda row: KMcost(row, 'VEHTYPE_B', 'NEW_Distance_B'),
    axis=1
)
# Second case
sli_df['B-carry-A-After_Colla_A'] = sli_df.apply(
    lambda row: KMcost(row, 'VEHTYPE_A', 'NEW_Distance_A'),
    axis=1
)
sli_df['B-carry-A-After_Colla_B'] = sli_df.apply(
    lambda row: KMcost(row, 'VEHTYPE_B', 'Prev_Dist_B') +
                KMcost(row, 'VEHTYPE_B', 'ORIG_DISTANCE') +
                KMcost(row, 'VEHTYPE_B', 'DISTANCE_A') +
                KMcost(row, 'VEHTYPE_B', 'DEST_DISTANCE') ,
    axis=1
)

# Calculate After_Colla
sli_df['After_Colla_A'] = (sli_df['A-carry-B-After_Colla_A'] + sli_df['B
```

```
    -carry-A-After_Colla_A'])/2
sli_df['After_Colla_B'] = (sli_df['A-carry-B-After_Colla_B'] + sli_df['B
    -carry-A-After_Colla_B'])/2


# Then, update the calculated results back to Cost_df
Cost_df.loc[sli_df.index, 'KMCost_Before_A'] = sli_df['Before_Colla_A']
Cost_df.loc[sli_df.index, 'KMCost_Before_B'] = sli_df['Before_Colla_B']
# Then, update the calculated results back to Cost_df
Cost_df.loc[sli_df.index, 'KMCost_After_A'] = sli_df['After_Colla_A']
Cost_df.loc[sli_df.index, 'KMCost_After_B'] = sli_df['After_Colla_B']


Cost_df.loc[sli_df.index, 'KMCost-A-carry-B-After_Colla_A'] = sli_df['A-
    carry-B-After_Colla_A']
Cost_df.loc[sli_df.index, 'KMCost-B-carry-A-After_Colla_A'] = sli_df['B-
    carry-A-After_Colla_A']

Cost_df.loc[sli_df.index, 'KMCost-A-carry-B-After_Colla_B'] = sli_df['A-
    carry-B-After_Colla_B']
Cost_df.loc[sli_df.index, 'KMCost-B-carry-A-After_Colla_B'] = sli_df['B-
    carry-A-After_Colla_B']

#none_approved
sli_df = Cost_df[(Cost_df['Collaboration_Status'] == 'none_approved')]
# Calculate Before_Colla
# Calculate energy consumption, directly sum the results, as NaN is
    already handled internally
sli_df['Before_Colla_A'] = sli_df.apply(
    lambda row: KMcost(row, 'VEHTYPE_A', 'Prev_Dist_A') +
                KMcost(row, 'VEHTYPE_A', 'DISTANCE_A'),
    axis=1
)
sli_df['Before_Colla_B'] = sli_df.apply(
    lambda row: KMcost(row, 'VEHTYPE_B', 'Prev_Dist_B') +
                KMcost(row, 'VEHTYPE_B', 'DISTANCE_B'),
    axis=1
)
# Calculate After_Colla_A
sli_df['After_Colla_A'] = sli_df.apply(
    lambda row: KMcost(row, 'VEHTYPE_A', 'Prev_Dist_A') +
                KMcost(row, 'VEHTYPE_A', 'DISTANCE_A'),
    axis=1
)
sli_df['After_Colla_B'] = sli_df.apply(
    lambda row: KMcost(row, 'VEHTYPE_B', 'Prev_Dist_B') +
                KMcost(row, 'VEHTYPE_B', 'DISTANCE_B'),
    axis=1
)
# Then, update the calculated results back to Cost_df
Cost_df.loc[sli_df.index, 'KMCost_Before_A'] = sli_df['Before_Colla_A']
Cost_df.loc[sli_df.index, 'KMCost_Before_B'] = sli_df['Before_Colla_B']
# Then, update the calculated results back to Cost_df
Cost_df.loc[sli_df.index, 'KMCost_After_A'] = sli_df['After_Colla_A']
Cost_df.loc[sli_df.index, 'KMCost_After_B'] = sli_df['After_Colla_B']
```

```python


# Calculate the final benefits
Cost_df["KM_Benefits_A"] = Cost_df["KMCost_Before_A"] - Cost_df["KMCost_After_A"]
Cost_df["KM_Benefits_B"] = Cost_df["KMCost_Before_B"] - Cost_df["KMCost_After_B"]

Cost_df["KM_Benefits_Total"] = Cost_df["KM_Benefits_A"] + Cost_df["KM_Benefits_B"]

#  Correct benefits_total values of 0 or negative to 'none_approved'

# Check the `KM_Benefits_Total` column, and if the value is 0 or negative, set `Collaboration_Status` to `none_approved`
Cost_df.loc[Cost_df['KM_Benefits_Total'] <= 0, 'Collaboration_Status'] = 'none_approved'

Cost_df['DEPTIME_Bias'] = (Cost_df['TRIP_DEPTIME_A'] - Cost_df['TRIP_DEPTIME_B'])/ Cost_df['TRIP_DEPTIME_A']


# Add a new column "Cost_savingrate"
Cost_df['Cost_savingrate'] = Cost_df['KM_Benefits_Total'] / (Cost_df['KMCost_Before_A'] + Cost_df['KMCost_Before_B'])
# Save the file
updated_file_path = "E:\\1MASTER in TUD\\@Master Thesis\\data\\range_considered\\range_0_KMCost.csv"
Cost_df.to_csv(updated_file_path, index=False)
```

4 Code for collaboration index research

```python
#%% Total cost for all trips in the data

import pandas as pd

# Load data
df = pd.read_csv("E:\\1MASTER in TUD\\@Master Thesis\\data\\preprocess_file_filter_capacity.csv")
ref = pd.read_csv("E:\\1MASTER in TUD\\@Master Thesis\\data\\Cost_VehType_2016.csv")

def KMcost(row, vehicle_type_col, distance_col):
    # Extract vehicle type and distance from the row
    vehicle_type = row[vehicle_type_col]

    distance = row[distance_col] if pd.notna(row[distance_col]) else 0

    # Find the cost per km for the given vehicle type in the ref dataframe
    cost_per_km = ref.at[vehicle_type, 'CostPerKm']
    # Calculate the total cost
    costKM = cost_per_km * distance / 1000
```

```python
    return costKM

# Use the apply function to calculate KMcost for each row, then sum all the results to get the total
total_km_cost_df = df.apply(lambda row: KMcost(row, 'VEHTYPE', 'Distance'), axis=1).sum()

print(f"Total Cost: {total_km_cost_df:.2f}")

#%%

import pandas as pd
import numpy as np

# Load data
file_path = "E:\\1MASTER in TUD\\@Master Thesis\\data\\range_considered\\range_1000_KMcost.csv"
Cost_df = pd.read_csv(file_path)

# Add a new column to calculate the total cost
Cost_df['Total_Before_Cost'] = Cost_df['KMCost_Before_A'] + Cost_df['KMCost_Before_B']
Cost_df = Cost_df[Cost_df['Collaboration_Status'] != 'none_approved']

# Parse the "PAIR_ID" column and split it into two new columns "Self_carrier" and "Participant_carrier"
Cost_df[['Self_carrier', 'Participant_carrier']] = Cost_df['PAIR_ID'].str.extract(r'\((\d+),\s*(\d+)\)')

# Convert the new columns to integers
Cost_df['Self_carrier'] = Cost_df['Self_carrier'].astype(int)
Cost_df['Participant_carrier'] = Cost_df['Participant_carrier'].astype(int)

# Create a 316x316 matrix with index and column names from 0 to 315
benefits_matrix = pd.DataFrame(np.zeros((316, 316)), index=range(316), columns=range(316))

# Create an array to store the TC_saving_rate for each carrier
TC_saving_rates = np.zeros(316)

# Set random sampling ratio and cost-saving rate threshold
sample_fraction = 1.0  # e.g., 30%
cost_saving_threshold = 0.2  # e.g., 10%

# **Initialize a dictionary to store the final DataFrame for each carrier
final_dfs = {}  # **

# **Initialize variables for total benefits and total cost
total_benefits = 0  #
total_before_cost = 0  #

# Iterate over all possible carrier_ids from 0 to 315
for self_carrier_id in range(316):
```

```
    # Filter rows where Self_carrier or Participant_carrier is the current self_carrier_id
    filtered_df = Cost_df[(Cost_df['Self_carrier'] == self_carrier_id) | (Cost_df['Participant_carrier'] == self_carrier_id)]

    # Randomly sample data
    sampled_df = filtered_df.sample(frac=sample_fraction, random_state=42)

    # Determine the primary carrier for consistent handling of carrier appearances
    sampled_df['carrier'] = self_carrier_id
    sampled_df['other_carrier'] = sampled_df.apply(lambda x: x['Self_carrier'] if x['Participant_carrier'] == self_carrier_id else x['Participant_carrier'], axis=1)

    # Calculate the occurrence count of each other_carrier with self_carrier
    participant_counts = sampled_df['other_carrier'].value_counts().to_dict()
    sampled_df['Participant_count'] = sampled_df['other_carrier'].map(participant_counts)

    # Add a new column to calculate the absolute value of DEPTIME_Bias
    sampled_df['DEPTIME_Bias_Abs'] = sampled_df['DEPTIME_Bias'].abs()

    # Sort according to the given rules
    sorted_df = sampled_df.sort_values(by=['KM_Benefits_Total', 'Participant_count', 'KMCost_After_A', 'Capacity_B', 'DEPTIME_Bias_Abs'],
                                       ascending=[False, False, True, True, True])

    # Ensure TRIP_ID_A and TRIP_ID_B each appear only once
    unique_trip_id_a_df = sorted_df.drop_duplicates(subset='TRIP_ID_A')
    final_df = unique_trip_id_a_df.drop_duplicates(subset='TRIP_ID_B')

    # Filter rows where 'Cost_savingrate' is greater than the specified threshold
    high_savings_df = final_df[final_df['Cost_savingrate'] > cost_saving_threshold]

    # **Save the final_df for each carrier
    final_dfs[self_carrier_id] = high_savings_df.copy()

    # Calculate total savings rate
    before_cost = high_savings_df['Total_Before_Cost'].sum()
    if before_cost > 0:
        TC_saving_rate = high_savings_df['KM_Benefits_Total'].sum() / before_cost
    else:
        TC_saving_rate = 0  # Or you can set it to np.nan, depending on your requirements
    TC_saving_rates[self_carrier_id] = TC_saving_rate

```

```python
    # *Accumulate total benefits and total cost
    total_benefits += high_savings_df['KM_Benefits_Total'].sum()  # *
    total_before_cost += high_savings_df['Total_Before_Cost'].sum()  # *

    # Aggregate calculation: For each other_carrier, calculate the total KM_Benefits_Total for transactions with self_carrier_id
    grouped_data = high_savings_df.groupby('other_carrier')['KM_Benefits_Total'].sum()

    # Populate the matrix
    for carrier_id, total_km_benefits in grouped_data.items():
        # Only populate if other_carrier is valid and within the range of 0 to 315
        if 0 <= carrier_id < 316:
            benefits_matrix.at[self_carrier_id, carrier_id] = total_km_benefits

print(f"Random selection rate: {sample_fraction}")
print(f"Cost-saving threshold: {cost_saving_threshold}")

# *Calculate the overall cost saving rate for all carriers
overall_cost_saving_rate = total_benefits / total_before_cost if total_before_cost > 0 else 0


def find_and_store_common_data(carrier_id_1, carrier_id_2, final_dfs):
    # Retrieve the final_df for two carriers
    df1 = final_dfs[carrier_id_1]
    df2 = final_dfs[carrier_id_2]

    # Find common indices between the two final_dfs
    common_indices = df1.index.intersection(df2.index)

    # If common indices exist, extract the data for these indices
    if not common_indices.empty:
        common_data = df1.loc[common_indices]

        # Calculate the proportion of common indices in each carrier's total indices
        proportion_1 = len(common_indices) / len(df1) if len(df1) > 0 else 0
        proportion_2 = len(common_indices) / len(df2) if len(df2) > 0 else 0

        # Return this data for further processing outside the function
        return common_data, proportion_1, proportion_2
    else:
        return None, 0, 0

# Initialize a dictionary to store all common data for each carrier_id
all_common_data_dict = {i: pd.DataFrame() for i in range(316)}

# Initialize a new matrix
new_benefits_matrix = pd.DataFrame(np.zeros((316, 316)), index=range(316), columns=range(316))
```

```

# Nested loop to iterate over all combinations
for i in range(316):
    for j in range(i+1, 316):  # Start from i+1 to avoid duplicate calculations and self-comparison
        common_data, prop_1, prop_2 = find_and_store_common_data(i, j, final_dfs)
        if common_data is not None:
            # Add the common data to the DataFrame related to i and j
            all_common_data_dict[i] = pd.concat([all_common_data_dict[i], common_data])
            all_common_data_dict[j] = pd.concat([all_common_data_dict[j], common_data])

            # Update the new matrix, recording the total KM_Benefits_Total of the common data in the corresponding position of the matrix
            new_benefits_matrix.at[i, j] = common_data['KM_Benefits_Total'].sum()
            new_benefits_matrix.at[j, i] = common_data['KM_Benefits_Total'].sum()  # Ensure matrix symmetry


# Initialize an array to store the cost_saving_rate for each carrier_id
TC_saving_newrates = np.zeros(316)

# Initialize total benefits and total cost
t_benefits = 0
t_before_cost = 0

# Iterate over all common data for each carrier_id, accumulating total benefits and total cost, and calculating the cost_saving_rate for each carrier
for carrier_id, df in all_common_data_dict.items():
    if not df.empty:
        benefits = df['KM_Benefits_Total'].sum()
        before_cost = df['Total_Before_Cost'].sum()
        # Accumulate to the total value
        t_benefits += benefits
        t_before_cost += before_cost
        # Calculate the cost_saving_rate for a single carrier
        cost_saving_rate = benefits / before_cost if before_cost > 0 else 0
        TC_saving_newrates[carrier_id] = cost_saving_rate

# Calculate the overall cost_saving_rate
overall_rate = t_benefits / t_before_cost if t_before_cost > 0 else 0

# Output results

print(f"Total Benefits (exclude option conflict) {t_benefits:.2f}")

#%%
import pickle

```

```
# Save dictionary data to a Pickle file
with open("E:\\1MASTER in TUD\\@Master Thesis\\data\\Suggestion-Plan.pkl
    ", 'wb') as pickle_file:
     pickle.dump(all_common_data_dict, pickle_file)

#%% Save the CI matrix
# Save the DataFrame as an Excel file
file_path = "E:\\1MASTER in TUD\\@Master Thesis\\data\\benefits_matrix.
    xlsx"
new_benefits_matrix.to_excel(file_path, index=True)

#%% Visualize the benefits matrix

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Ensure the path is correct
file_path = "E:\\1MASTER in TUD\\@Master Thesis\\data\\benefits_matrix.
    xlsx"
CI = pd.read_excel(file_path, sheet_name='Sheet2', index_col=0)

# Set up the matplotlib figure
plt.figure(figsize=(8, 6))

# Draw the heatmap with the "plasma" colormap
sns.heatmap(CI, cmap="plasma")

# Add title and labels
plt.title("Heatmap of the Collaboration Index Among Carriers with IDs
    Ranging from 20 to 30")
plt.xlabel('Carrier ID')
plt.ylabel('Carrier ID')

# Show the plot
plt.show()

# Save the plot with high resolution
plt.savefig("E:\\1MASTER in TUD\\@Master Thesis\\methodology\\CI-heatmap
    .png", dpi=300)

#%% CI matrix final display code
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np

# Ensure the path is correct
file_path = "E:\\1MASTER in TUD\\@Master Thesis\\data\\benefits_matrix.
    xlsx"
CI = pd.read_excel(file_path, sheet_name='Sheet2', index_col=0)

# Create a mask for the upper triangle without including the diagonal
mask = np.triu(np.ones_like(CI, dtype=bool), k=1)

```

```
# Set up the matplotlib figure
plt.figure(figsize=(12, 10))

# Draw the heatmap with the "plasma" colormap, applying the mask and
    formatting the annotations
sns.heatmap(CI, cmap="plasma", mask=~mask, annot=True, fmt=".0f",
    annot_kws={"size": 12},cbar_kws={"shrink": .8})

# Add title and labels
plt.title("Heatmap of the Collaboration Index Among Carriers with IDs
    Ranging from 20 to 30")
plt.xlabel('Carrier ID')
plt.ylabel('Carrier ID')

# Save the plot with high resolution
plt.savefig("E:\\1MASTER in TUD\\@Master Thesis\\methodology\\CI-heatmap
    .png", dpi=500)

# Show the plot
plt.show()

#%% Calculate cost savings for the optimized routes compared to the
    initial total route cost for each carrier
import pickle

# Load the dictionary data from the Pickle file
with open("E:\\1MASTER in TUD\\@Master Thesis\\data\\Suggestion-Plan.pkl
    ", 'rb') as pickle_file:
     a_dict_plan = pickle.load(pickle_file)

import pandas as pd

# Iterate over all keys in the dictionary
for key in a_dict_plan.keys():
    df = a_dict_plan[key]  # Get the dataframe for the current key

     # Ensure the dataframe has the required columns
     if 'carrier' in df.columns and 'other_carrier' in df.columns:
         # Create a new column 'AorB' based on the comparison
         df['AorB'] = df.apply(
             lambda row: 'True' if row['carrier'] == key
             else 'False' if row['other_carrier'] == key
             else None, axis=1)

         # Check for rows where neither condition is met and print those
    rows
         for idx, row in df.iterrows():
             if row['AorB'] is None:
                 print(f"Key: {key}, Row: {idx}, Data: {row.to_dict()}")
     else:
         print(f"Key {key} does not have the required columns 'carrier'
    and 'other_carrier'.")

# Update the dictionary with the modified dataframes
a_dict_plan_updated = a_dict_plan
```

```python

#%% Calculate cost savings for each company when collaborating with another company based on its own trip cost reduction
import numpy as np
import pandas as pd
import pickle

# Number of carriers
num_carriers = 316

# Initialize the cost savings matrix
cost_savings_matrix = np.zeros((num_carriers, num_carriers))

# Iterate over all keys in the dictionary
for key in a_dict_plan_updated.keys():
    df = a_dict_plan_updated[key]  # Get the dataframe for the current key

    # Ensure the dataframe has the required columns
    if 'AorB' in df.columns and 'KM_Benefits_A' in df.columns and 'KM_Benefits_B' in df.columns:
        # Calculate the cost savings for the current carrier
        for idx, row in df.iterrows():
            if row['AorB'] == 'True' and row['KM_Benefits_A'] > 0:
                other_carrier = row['other_carrier']
                cost_savings_matrix[key, other_carrier] += row['KM_Benefits_A']
            elif row['AorB'] == 'False' and row['KM_Benefits_B'] > 0:
                other_carrier = row['carrier']
                cost_savings_matrix[key, other_carrier] += row['KM_Benefits_B']

# Create a DataFrame from the cost savings matrix
cost_savings_own = pd.DataFrame(cost_savings_matrix)

# Define the file path to save the Excel file
file_path = "E:\\1MASTER in TUD\\@Master Thesis\\data\\ownsaving_matrix.xlsx"

# Save the DataFrame to an Excel file
cost_savings_own.to_excel(file_path, sheet_name='Own_saving')

#%% Calculate the total trip cost for each carrier
import pandas as pd

# Load data
df = pd.read_csv("E:\\1MASTER in TUD\\@Master Thesis\\data\\preprocess_file_filter_capacity.csv")
ref = pd.read_csv("E:\\1MASTER in TUD\\@Master Thesis\\data\\Cost_VehType_2016.csv")

def KMcost(row, vehicle_type_col, distance_col):
    # Extract vehicle type and distance from the row
    vehicle_type = row[vehicle_type_col]

```

```
    distance = row[distance_col] if pd.notna(row[distance_col]) else 0

    # Find the cost per km for the given vehicle type in the ref dataframe
    cost_per_km = ref.at[vehicle_type, 'CostPerKm']
    # Calculate the total cost
    costKM = cost_per_km * distance / 1000
    return costKM

# Calculate KMcost for each row and store it in a new column
df['KMcost'] = df.apply(lambda row: KMcost(row, 'VEHTYPE', 'Distance'), axis=1)

# Group by carrier and calculate the total cost
total_cost_per_carrier = df.groupby('CARRIER_ID')['KMcost'].sum()

# Print the total cost for each carrier
print(total_cost_per_carrier)

#%% Calculate cost savings percentage compared to the initial total route cost for each carrier
# Ensure the DataFrame has carrier IDs as the index
cost_savings_own.columns = cost_savings_own.columns.astype(int)  # Ensure columns are integers if they're carrier IDs

# Align the DataFrame and the Series by index for division
aligned_costs = cost_savings_own.div(total_cost_per_carrier, axis=1)

# Replace NaN values with zero if any exist after the division
aligned_costs.fillna(0, inplace=True)

# Calculate the total percentage savings per carrier
total_percentage_savings_per_carrier = aligned_costs.sum() * 100  # Calculate the sum of each column and convert to percentage

# Append the total savings as a new row in the DataFrame
aligned_costs.loc['Total Percentage Savings'] = total_percentage_savings_per_carrier

# Save the updated DataFrame to an Excel file
aligned_costs.to_excel("E:\\1MASTER in TUD\\@Master Thesis\\data\\save_rate_matrix_per_carrier.xlsx")

# Optionally, print the DataFrame to see the new row
print(aligned_costs)

#%% Boxplot visualization
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

file_path = "E:\\1MASTER in TUD\\@Master Thesis\\data\\save_rate_matrix_per_carrier.xlsx"
aligned_costs = pd.read_excel(file_path, sheet_name='Sheet1', index_col
```

```
    =0)

# Assume aligned_costs is a defined DataFrame
row_data = aligned_costs.loc['Total Percentage Savings']  # Assuming we
    have the actual data here

# Create a boxplot using seaborn
plt.figure(figsize=(12, 6))
box = sns.boxplot(data=row_data, orient='h', palette="muted", showmeans=
    True,
                  meanprops={"marker":"o", "markerfacecolor":"red", "
    markeredgecolor":"black"})

# Calculate quartiles and mean
q1 = np.percentile(row_data, 25)
q3 = np.percentile(row_data, 75)
mean = np.mean(row_data)
median = np.percentile(row_data, 50)

# Add annotations for quartiles and mean
plt.text(median, 0.5, f'Median: {median:.2f}', horizontalalignment='
    center', color='black', weight='semibold', fontsize=13)
plt.text(mean, 0.1, f'Mean: {mean:.2f}', horizontalalignment='center',
    color='red', weight='semibold', fontsize=13)
plt.text(q1, 0.5, f'Q1: {q1:.2f}', horizontalalignment='center', color='
    blue', weight='semibold', fontsize=13)
plt.text(q3, 0.5, f'Q3: {q3:.2f}', horizontalalignment='center', color='
    green', weight='semibold', fontsize=13)

# Set plot title and labels
plt.title('Box Plot of Cost-saving Rate Across Companies', fontsize=16)
plt.xlabel('Cost-saving Rate', fontsize=14)

# Show the plot
plt.show()

# Save the image as a high-resolution PNG file
plt.savefig('E:\\1MASTER in TUD\\@Master Thesis\\methodology\\
    cost_boxplot.png', dpi=500)
```

Appendix B

The scientific paper is presented in the following pages.

Developing a Collaboration Index for Carriers: Identifying Opportunities in Road Freight Transport

Tangzhe Chen
Civil Engineering
TU Delft
Delft, Netherlands

***Abstract*—In response to the inefficiencies in road freight transport, this research proposes an innovative collaboration index for carriers. The framework involves a collaboration decision tree that determines four collaboration modes at the trip level, which are then aggregated into a carrier-level evaluation using heuristic local optimization. This results in a collaboration index that quantitatively assesses the potential for collaboration among carriers based on operational data, filling a gap in partner selection processes. The study reveals that participants in horizontal collaboration achieved an average cost-saving rate of 14.4%, with the top 25% reaching 22.12%. While geographical similarity moderately correlates with the index, vehicle type does not show a linear relationship.**

The significance of this work lies in its unique collaboration index formulated from operational data, which not only facilitates the identification of collaboration potentials but also provides insights for collaborative partner selection strategies.

***Index Terms*—horizontal collaboration, freight transport, partner selection, collaboration index**

I. Introduction

A. Research context

In 2022, the European Union experienced a significant amount in road freight transport, with a total volume exceeding 13.6 billion tonnes and covering an extensive 1,920 billion tonne-kilometres. Road freight transport constantly increases nowadays. According to the European Union Statistics, the growth of total road freight transport developed at 2.1% every year from 2018 to 2022. A concerning statistic emerged, revealing that at the EU level, 20.2% of the total distances covered were by empty vehicles, posing a considerable challenge to the overall effectiveness. This issue was prevalent across most countries, with high percentages of empty vehicles in both national and international transport. In 2022, at the EU level, 20.2% of the total distances performed by empty vehicles. The Netherlands, in particular, stood out with a staggering 32.6% of empty vehicles in national transport and 22.4% in international transport, surpassing 20% which is mirrored by only a few countries (Eurostat, 2023).

One negative effect of certain inefficient logistics is the cost increase in operation, particularly highlighted by the existence of empty vehicles. This situation occurs when trucks and delivery vehicles travel without cargo, leading to higher operational costs as companies must cover the expenses for these non-revenue-generating trips (McKinnon & Ge, 2006). Moreover, these inefficiencies intensify traffic congestion, which can lead to higher infrastructure maintenance costs.

The shortage of freight transport efficiency not only increases the cost of travelling but leads to other negative consequences as well, for example, the negative impact on the environment of greenhouse gas (GHG) emissions. In 2021, road vehicles contributed about three-quarters of the European Union's transportation emissions. Petroleum-based fuel combustion in road transportation emerged as the largest source of CO_2 emissions in the EU. Passenger cars were responsible for 60% of road transportation CO_2 emissions, while heavy-duty vehicles (trucks and buses) ranked second, emitting over 200 million tonnes of carbon dioxide equivalent annually (200 Mt CO_2e), comprising about one-quarter of road transportation emissions (Statista, 2024). Although heavy-duty vehicles have fewer vehicles on the road, high levels of pollution are emitted by them, emphasizing the need to address their environmental problem. The reduced number of freight transport trips, especially the empty vehicles would help address this problem.

Collaborative logistics has been believed to be one innovative method for tackling the inefficiency in logistics (Amer & Eltawil, 2015; Cruijssen et al., 2007; Y. Wang et al., 2021). Collaborative logistics is a strategy in which companies join together to enhance the effectiveness of their supply chains. This approach contrasts with working isolate accompanied by inefficiency inherent (Ferrell et al., 2020). Horizontal and vertical collaboration are divided by different physical network structures. Horizontal collaboration involves the cooperation of multiple organizations that are either unrelated or competitors. They are at the same logistics level, cooperating for sharing individual information or assets, such as vehicle capacity. On the other hand, vertical collaboration is the joint effort of multiple organizations functioning at various levels of the supply chain. This could include for example manufacturers, distributors, shippers, and carriers, who collaborate by sharing their duties, resources, and data to cater to a similar group of end customers within a specific supply chain (Ferrell et al., 2020; Vargas et al., 2018).

Some research focuses on reviewing the existing literature and published case studies. It has been noted that studies focusing on collaborations often lack grounding in real-world data. Gansterer and Hartl (2020) suggested that the data analysis method such as neural neural network could be used for a deeper understanding of effective strategy applied in the collaboration environment. Key difficulties and obstacles for collaboration are highlighted theoretically including the distribution of revenue, competition law and policy, process coordination, organization compatibility, structure and strategies cooperation (Audy et al., 2012; Vargas et al., 2018). To better understand the nature of barriers, Karam et al. (2021) categorized 31 recognized obstacles into five groups, business models, information exchange, human factors, Collaborative Decision Support Systems (CDSSs), and market influences. The model helped to guide the risks and benefits of implementing collaborations.

Researchers categorized studies related to horizontal collaboration by the context of the problem. The classification includes partner selection and relationship management (Amer & Eltawil, 2015; Essam & Eltawil, 2014). The partner selection research tries to identify and test the strategic fit of companies before they collaborate. Additionally, within the scope of relationship management, issues are further divided into areas such as forecasting, sharing of profits/costs, and collaborative approaches for sharing capacity/demand and routing problems (Ergun et al., 2007; Lai et al., 2022; J. Wang et al., 2018; Y. Wang et al., 2021). The review of literature on the topics of partner selection indicates frequent use of the Analytic Hierarchy Process (AHP) method (Bahinipati et al., 2009; Feng et al., 2010; Naesens et al., 2009). This preference is attributed to the proficiency of AHP in organizing multi-attribute issues into a hierarchical structure, which allows for a detailed examination of each hierarchical level independently. The research on relationship management is much more than the partner selection (Essam & Eltawil, 2014).

The current research primarily focuses on broader perspectives, such as financial aspects, historical sales data, and employee counts. However, there is a noticeable gap in the literature regarding the selection of partners in the operational domain, specifically on trips and relevant freight information. Additionally, while there is a substantial amount of research dedicated to investigating decision-making processes, the step for addressing the identification of potential collaborations among companies through data analysis methods is relatively neglected.

In summary, this research aims to develop an identification collaboration index for road freight transport. The exploration of the collaboration index provides insights between carriers for collaboration.

B. Research objective

The objective of this research is to define a collaboration index to identify opportunities for collaboration among carriers. The collaboration index will be developed to assess the potential for collaboration among the companies represented in the dataset.

By developing the collaboration index, the research contributes to guiding companies and authorities for partner selection and evaluation when considering collaborative freight transportation. The quantitative insights induced by the index introduced a direct method considering various attributes and factors for collaboration. Additionally, as a foundation for further collaborative decision-making problem optimization, this research provides a quantitative way for data filtering and data prepossessing.

C. Research scope

Building on the concepts outlined in the preceding section, this research aims to address the gap in understanding horizontal collaboration within freight transportation, with a specific emphasis on partner selection. Consequently, the scope of this study is confined to the aspects of collaboration partner selection, and it will not delve into the optimization of various collaboration problems.

This research investigates under the assumption that all companies in the dataset are equal in terms of background, omitting factors such as financial status, number of employees, and other criteria typically considered in current partner selection studies. The focus of this study is exclusively on freight transportation data.

Also, the collaboration index developed in this research is applied at the company (carrier) level, signifying that it represents an aggregate measure based on the cumulative data of trips and tours by each carrier. This index serves as a tool to evaluate the extent of the collaborative potential of a carrier, taking into account the comprehensive logistic activities encompassed in their freight operations. By aggregating individual trips and tours, the index offers an overall insight into the collaboration potential of multiple companies in the dataset.

D. Main research question & sub-questions

Based on the objective introduced in section I-B, the main research question is defined as follows which addressed the objectives:

How can the 'collaboration index' be defined to assess the potential for collaboration among companies involved in road freight transport?

The main question is answered by the following sub-questions:

1) How can the 'collaboration index' be used by companies and authorities for partner selection and evaluation in freight collaboration transportation?
2) What attributes and factors should be included in the 'collaboration index' to reflect the potential for collaborative logistics accurately?
3) How can the 'collaboration index' be formulated to quantitatively measure the potential for collaboration?

II. LITERATURE REVIEW

This Section will comprehensively review and introduce relevant literature related to this research. By exploring scholarly works, theories, and studies, this section aims to establish a solid foundation, enriching the understanding of the topic and objective for answering the research question.

A. *Collaborative transport networks*

The Collaborative Transport Networks (CTNs) represent an innovative approach to logistics and supply chain management, where a group of competing companies, including shippers and carriers, opt to collaborate in their freight transport operations (Pan et al., 2019). This collaboration aims to design and implement a shared logistics and transport network, optimizing efficiency and resource utilization across the supply chain. By pooling resources and coordinating efforts, CTNs can achieve significant benefits, including lower transportation expenses, better service quality, and improved sustainability. The approach of CTNs fosters a collaborative relationship among partners, transforming competitive dynamics into cooperative strategies to tackle common logistical challenges effectively (Karam et al., 2021).

In Collaborative Transport Networks (CTNs), the companies (partners) have to collaborate by sharing information about their transport orders and available trucks with a central coordinator. This coordinator uses the shared data to identify opportunities for collaboration. The coordinator, acting as a neutral third party, is trusted by the collaborating partners to manage the collaboration impartially and ensure the secure handling of shared information (Audy et al., 2007).

To handle the large number of information shared from the partners, the coordinator need a system for processing. Collaborative Decision Support Systems (CDSSs) are implemented to manage and facilitate these collaborative processes (Gansterer & Hartl, 2018).

1) Collaborative Decision Support Systems: Collaborative Transport Networks (CTNs) rely on Collaborative Decision Support Systems (CDSSs) for information sharing and the development of collaborative plans. The CDSS typically comprises three main components: a database, computational algorithms, and an interactive dialogue. The overview of the CDSS is shown in Figure 1. The database module collects and retains all information related to logistics, which can be input manually or automatically from partners' transport planning systems. This data is then utilized by the algorithmic module, which analyzes the shared data and makes collaborative decisions, including proposals for exchanges and the distribution of profits among partners (Karam et al., 2021). Compared to non-collaborative efforts, the financial advantage gained from collaboration is termed the benefits of collaboration or profits. CDSSs are frequently employed to allocate these profits among the participating companies using various profit-sharing mechanisms (Guajardo & Rönnqvist, 2016).

There are two main types of algorithms, which are order sharing and capacity sharing. Order sharing requires the order detail information sharing from all the participants including the customer orders with product flows. A central coordinator would optimize orders, and assign the orders matching with participants (Dahl & Derigs, 2011). For the capacity sharing, it requires the sharing of available capacities within the trucks, excluding sensitive customer order information. Capacity sharing is also more applied in companies when considering collaboration (Agarwal & Ergun, 2010).

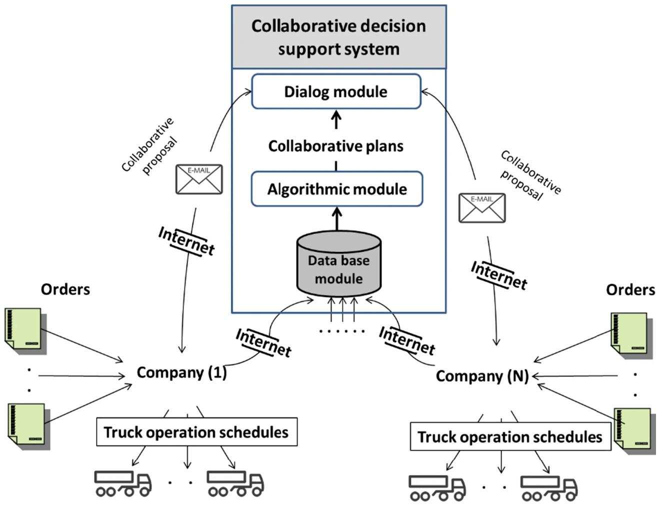


Fig. 1. Overview of Collaborative Decision Support System (Karam et al., 2021)

When considering the partners involved in CDSSs, the user might be carriers, shippers or receivers. The single carrier collaboration represents a peer-to-peer bilateral partnership between carriers, driven by objectives such as reducing transportation costs and enhancing service levels (Pan et al., 2019).

Some research addressed the collaboration partner selection by developing new approaches. Feng et al. (2010) outlined both individual and collaborative attributes and their respective measurements. They proposed a fuzzy multiple attribute decision-making (FMADM) approach to combine assessment data from both sets of utilities, facilitating the comprehensive ranking of potential partners. Awasthi et al. (2016) introduced a pioneer study with fuzzy BOCR-GRA methodology. Through a multi-step process involving criteria identification via BOCR, linguistic assessment with fuzzy logic, and ranking with GRA, this work innovatively navigated the complexities of partner selection under data constraints. The application of a veto mechanism and sensitivity analysis ensured robust and unbiased decision-making, marking a significant advancement in urban logistics planning.

Guajardo and Rönnqvist (2016) investigated methods for sharing the benefits of collaboration, identifying various industrial cases and the resulting cost savings. Their findings indicate that collaboration can yield cost savings ranging from 5.99% to 46%. Similarly, Hernández and Peeta (2014) introduced the concept of a degree of collaboration in their model for centralized multiple-carrier collaboration. A higher degree signifies a greater extent of collaboration among carriers, leading to cost savings between 2% and 55%. Recent research by Beliën et al. (2017) demonstrated that horizontal collaboration can reduce transportation costs by 10-15%.

Given the advancements in research methodologies for selecting collaboration partners, as highlighted by Feng et al.

(2010) and Awasthi et al. (2016), methodologies need proper factors related being included. The formulation of a collaboration index necessitates the consideration of several factors related, so the review of relevant literature is important. While there are no directly comparable studies currently available for partner selection, exploring optimization papers and topics that are closely related is crucial for grounding our research in a solid theoretical framework.

Main elements were identified and introduced through the classification of potential benefits and barriers (Karam et al., 2021). Information sharing, business models, Collaborative Decision Support Systems (CDSS), market dynamics, and human factors constitute the five primary elements of collaboration as introduced. Among the potential benefits derived from these elements are improved service levels, reduced empty runs, decreased traffic congestion, shorter delivery times, higher resource utilization, reduced lead times, and lower transportation costs. These benefits contribute guide considering various influencing factors to build effective collaborations.

In assessing the optimization model for collaborative logistics, Ghomi et al. (2023) incorporated truck types as a factor. The model accounted for the volume and weight of goods transported by various trucks, as well as the unused weight and volume capacities within these trucks. This analysis differentiated Truck-Load (TL) and Less-than-Truck-Load (LTL) trucks. To evaluate the effectiveness of different collaborative setups, the model considered the average numbers of these truck types and their respective transport distances.

In exploring how logistics platforms can enhance the utilization of vehicle capacity, a study into these platforms reveals key elements in collaboration. Deng et al. (2023) considered the creation of a three-player evolutionary game model (TEGM), which was instrumental in examining the dynamics among participants under various situations. This model's parameters were categorized in alignment with the stakeholders involved. Specifically, for freight carriers, the model takes into account aspects like the likelihood of adopting a shared capacity strategy, associated sharing expenses, payments, potential economic losses, and the resulting income for carriers when different strategies are employed.

B. Conclusion

Following a comprehensive review of existing scientific literature, several notable gaps in knowledge have been identified. Firstly, while Collaborative Transportation Networks (CTNs) offer an innovative approach for managing collaboration, current research predominantly focuses on relationship management within the decision-making process, with less emphasis on partner selection. Secondly, within partner selection, existing studies focused on qualitative aspects such as financial, historical sales data, employee counts, and goal motivations. However, there has been less emphasis on the operational domain, particularly analyses based on freight logistics data. This gap highlights the need for utilizing operational data to identify and optimize collaborative opportunities in freight transportation. Thirdly, the concept of a collaboration index represents an innovative method for quantitatively determining the potential for collaboration between companies, which can be included in the Collaborative Decision Support Systems (CDSS) as an evaluation value to highlight collaboration opportunities. Stakeholders such as coordinators, who act as neutral third parties, maintain the information and offer opportunities to participants. Customers in the CTN engage in the identification process for collaboration, seeking suitable partners. Additionally, the government or authorities may utilize the collaboration index as an indicator to assess the level of collaboration among candidates.

These identified gaps serve as the fundamental for the research objectives of this paper. Table I provides a comprehensive overview of the distinctions between this research and the relevant literature.

Accordingly, this study aims to delve into the collaboration using freight transport data, which considers several attributes associated with tours and trips. Through the application of the introduced approach, this research seeks to develop an effective 'collaboration index'. This index will be used as an indicator to assess the potential for collaboration. Subsequently, the collaboration index will be utilized to identify suitable partners and evaluate their compatibility for freight transportation, facilitating contribution to the field of logistics and supply chain management.

TABLE I
COMPARISON ANALYSIS BETWEEN THE CURRENT RESEARCH AND LITERATURE

Literature	Main Topic	Type of Collaboration	Methodology
Amer and Eltawil (2015)	Green Collaborative	Partner selection Relationship management	Literature research
Dahl and Derigs (2011)	Cost/profit allocation	Relationship management	Simulation
Y. Wang et al. (2021)	Vehicle routing problem	Relationship management	2E-CMDPVRP
Awasthi et al. (2016)	City logistics planning	Partner selection	Fuzzy BOCR–GRA
Feng et al. (2010)	Codevelopment alliance	Partner selection	FMADM
Bahinipati et al. (2009)	Collaboration evaluation	Partner selection	AHP-FLM
Luan et al. (2022)	Opportunity identification	Partner selection	Trailer capacity graph
Hernández and Peeta (2014)	Multihub Location Problem	Optimization	CCCMLP
This research	**Collaboration index identification**	**Partner selection**	**Quantitative design**

III. RESEARCH METHODOLOGY

In this Section, the research methodology is divided into two main sections: the determination of collaboration and the formulation of the collaboration index. The first section introduces a comprehensive decision tree designed to identify and classify four distinct modes of horizontal collaborations at the trip level. This detailed flowchart serves as a systematic tool to distinguish various collaboration scenarios, ensuring a clear categorization.

The second section builds upon the identified collaboration modes, delving deeper into the evaluation and quantification processes. This involves transforming the insight of trip-level collaboration into carrier-level collaboration information. By analyzing the trip-level data, the comprehensive collaboration index that reflects the extent and potential of collaborations among carriers is introduced. This index facilitates a better understanding of efficient collaboration and provides a metric for showing the collaborative potential between partnerships in freight road transportation carriers.

A. Collaboration determination

Before starting the investigation into collaboration, it is important to note that extensive literature has provided various complex methods for exploring horizontal collaboration optimization among candidates. However, there is a notable gap in research focusing on the partner selection step, specifically in suggesting which pairs of partners should collaborate. Additionally, due to the unavailability of complete customer data for each carrier, particularly detailed production-consumption data for cargo, this section develops a straightforward decision tree to identify four different modes of collaboration using available freight transportation data.

1) Assumptions: Before introducing the decision tree, it is essential to outline the specific assumptions made in this study. Firstly, the detailed cargo flow from origin to destination is not considered, allowing for a focus on more generalized aspects of the weight movement. Additionally, this research does not take into account the volume used by the cargo of the vehicles; instead, it primarily considers the weight of the cargo. It is assumed that for independent freight transportation (no collaboration), all cargo carried in a single vehicle during its tour is homogeneous, except for variations in weight that occur during loading and unloading at different nodes. These weight variations, observed before and after operations at each node, indicate whether the vehicle is engaged in either a loading or unloading service at that particular node. This assumption simplifies the complexity of cargo management and focuses the analysis on the critical aspect of weight management.

2) Collaboration decision tree: This section outlines a methodology for identifying four collaboration modes at the trip level using a decision tree. This approach examines the conditions under which two trips from different carriers can potentially collaborate in the context of freight truck transportation under the assumptions introduced in Section III-A1. Key variables considered include the routes of the trucks, their capacities, loads, trip weights, and types of cargo. The decision tree provides a logical framework to determine whether trip pairs can collaborate, thereby classifying the collaboration modes.

The decision tree for trip-level collaboration determination is shown in Figure 2. The explanation of each part of the decision tree is introduced as follows.

The initial step involves selecting two trips from two different carriers from the freight truck database. The decision tree for collaboration determination focuses on the trip level.

The analysis aims to distinguish whether the origin and destination (OD) nodes of the trip pairs fall within an acceptable range to facilitate carrier reallocation. Specifically, if both carriers' ORIG pair and DEST pair nodes are located within a predetermined range, it suggests a strong potential for collaboration. This approach enables more flexible and efficient coordination without strict adherence to fixed trip routes. The two carrier being determined in the decision tree named as carrier A and carrier B.

Determining the compatible cargo type involves assessing whether different types of cargo can be safely and efficiently stored in the same vehicle. This evaluation ensures that the combined transportation of various cargo types does not compromise the safety or quality of the goods being transported. Trip pairs that do not meet the cargo compatibility criteria are excluded from further consideration.

The next step addresses the scenarios where a carrier's depot coincides with the "origin" in the trip (OD) pairing. Depots serve as critical locations for carriers to prepare cargo and maintain their vehicles, necessitating their inclusion in the route of each tour. In cases where the "origin" is a depot for one of the carriers, these cases require detailed examination.

A notable exception arises when a trip involves a depot as the origin and only two nodes are involved along the whole tour. This exception occurs if the cargo can be transported by the other carrier through collaboration. Then the trip involving the depot can be saved and eliminated. Otherwise, collaboration is not feasible under these conditions.

The next step involves analyzing key decision branches within the collaboration process, categorized into three scenarios based on whether the carriers are unloading or loading at the "Origin". Each carrier has its own transport task, which involves transporting cargo between origin-destination (OD) trips. The transport task can be shared between two carriers only if the cargo is loaded at the "Origin" node, which allows the other carrier to also carry the cargo on its vehicle. Only the transport tasks that involve loading at the "Origin" are available for sharing. These scenarios are both carriers unloading, both carriers loading, and one carrier unloading while the other is loading.

In the scenario where both carriers are unloading, collaboration is not feasible, as the transport tasks for both carrier are not loading cargo from "Origin" to "Destination". Conversely, in the mixed scenario of one is loading and the other one is unloading, the carrier that is unloading has the potential to handle the transport task belong to the other carrier only if that does not exceed its capacity. If this condition is met, collaboration is initiated, with the unloading carrier responsible for the transport task originally belonging to the other carrier. If not, the collaboration attempt fails.

For the scenario where both carriers are loading, two inequalities need to be assessed: Inequality 1 determines if carrier A can carry cargo used to be responsible by carrier B at "Origin", while Inequality 2 checks if carrier B can carry the cargo used to be responsible by carrier A. Inequality 1 and Inequality 2 represent two different types of collaboration. Based on the satisfaction of these inequalities, four possible outcomes are listed.

A total of four collaboration modes are proposed: "A carry cargo for B" indicates that carrier A and B can collaborate by sharing carrier A's capacity with carrier B; "B carry cargo for A" signifies the reverse. "both approved" means both carriers can share their capacity with each other, while "no collaborate"

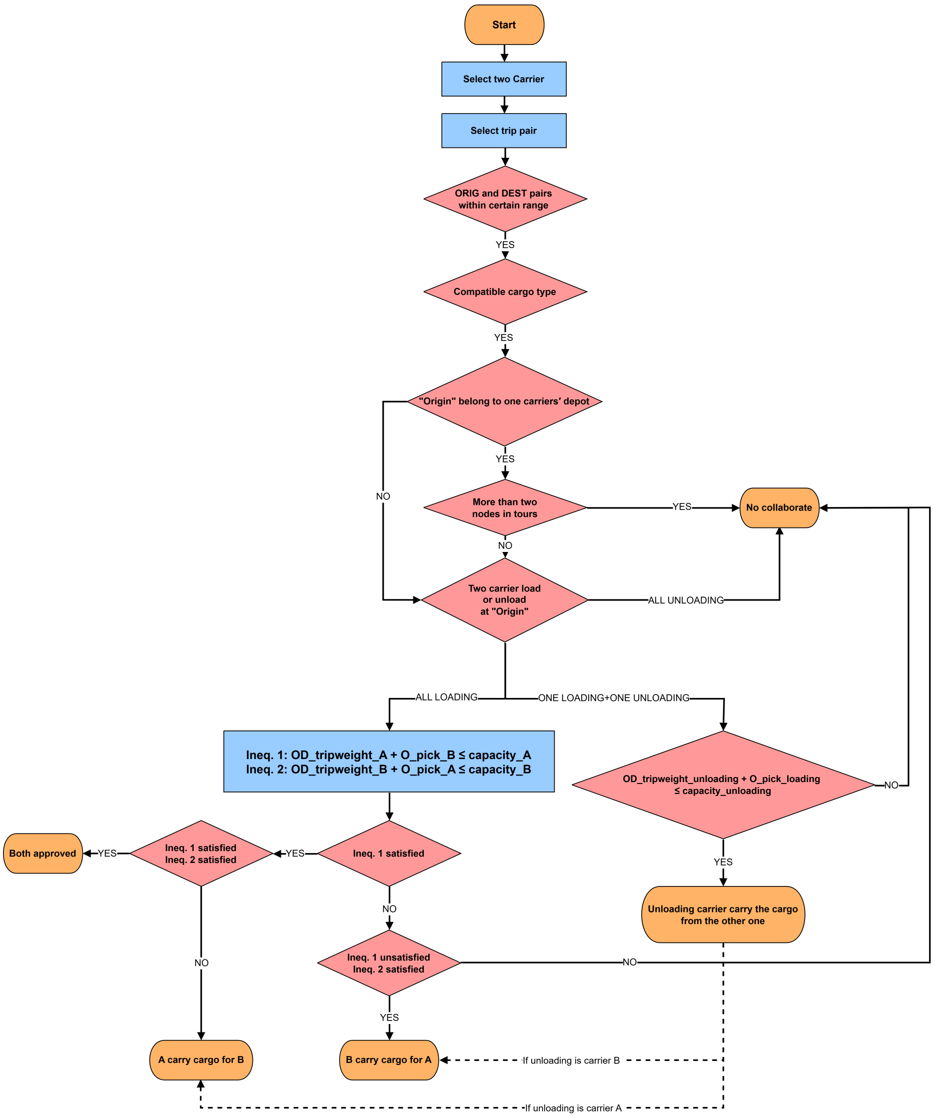


Fig. 2. Decision tree for trip-level collaboration determination

indicates that carriers A and B cannot collaborate for the chosen trips.

This decision tree outlines how two carrier trip pairs might collaborate based on detailed freight transportation data. The output modes identify the type of collaboration for different trip pairs, setting the stage for further research that will incorporate quantitative analysis to evaluate the performance of these collaborations.

B. Collaboration index formulation

The collaboration decision tree identifies four collaboration modes at the trip level. Building on these modes, this section employs a heuristic local optimization approach to evaluate collaboration at the carrier level. This involves analyzing how trip-level collaborations translate to broader carrier-level partnerships. Subsequently, a collaboration index is formulated at the carrier level, which quantifies the collaboration potential between different carriers. This index serves as a comprehensive metric for assessing carrier collaborations in freight transportation.

1) Heuristic local optimization: The collaboration determination provides collaboration options with four modes (three of which are feasible for collaboration), and an optimization approach needs to be applied to choose the relative best collaboration options between carriers. Given the lack of complete data, particularly cargo origin/destination (O/D) data, a generalized alternative to optimization is considered to develop simple decision rules that are easy to apply in practice. Heuristic local optimization is applied based on the trip-level results of the collaboration identification using the decision tree in Section III-A2. The rules of heuristic local optimization are constructed based on the main elements of collaboration found in the literature (Karam et al., 2021) to sort the collaboration options. This method provides a local optimization approach and can be compared to optimization models or data-driven approaches in future studies.

A straightforward method for evaluating collaboration is cost reduction. The predefined cost per kilometre parameters for seven different types of vehicles transforms the distance savings (km) achieved through collaboration into monetary cost reduction (€). The cost savings from the collaboration are treated as benefits. These benefits, quantified in monetary terms, directly reflect the savings each carrier achieves through collaboration. The mode "both approved" indicates that the collaboration plans "A carry cargo for B" and "B carry cargo for A" are both feasible for the trip pairs. This mode is unique compared to other modes in that it calculates the benefits for both scenarios and uses their average as the final benefit value. For the "no collaborate" mode, the benefits are set to zero.

The trip-level collaboration determination identifies potential collaboration options for carriers, but the trip pairings are not necessarily unique. For a specific carrier, a single trip may present multiple collaboration opportunities. Instead of solving for the highest global profit through a complex optimization problem, several metrics are defined and applied to sort these trip-level collaboration options, receiving a local optimal for each carrier. These metrics reflect the reference introducing the potential benefits of horizontal collaboration, including lower transport costs, less empty running, less carbon emissions, less traffic congestion, shorter delivery times, higher resource utilization, etc (Karam et al., 2021).

These metrics are considered one by one in the following order:

1) Total Profit: Prioritize collaboration options that offer the highest positive total profit.
2) Frequency of Collaboration: Prefer participant carriers that have more frequency options with Carrier A, as these carriers may indicate more stable and reliable relationships.
3) Tour Savings: Select the option that saves an entire tour for Carrier A, implying fewer vehicles needed and thereby reducing costs.
4) Vehicle Capacity of Participant: Opt for participant vehicles with lower capacity since smaller vehicles generally incur lower costs, enhancing flexibility and efficient resource use in the collaboration.
5) Departure Time Differences: Choose options with the smallest differences in departure times with trip in carrier A to minimize waiting times and potential scheduling conflicts.

Viable options are then selected in order, ensuring that each trip from a carrier is paired only once with the most suitable trip from a participating carrier. This method ensures that each trip pair appears uniquely in the selection process, preventing redundancy and repeated pairings in decision-making and analysis.

Based on the five metrics building for the sorting metrics, and the unique selection, each carrier identifies its own optimal collaborative plan. Although these plans give suggestions for each carrier, showing the optimal collaborative plan from their individual perspective, the plans do not lead to a globally optimal solution. This is because conflicts can arise between the plans of different carriers. For instance, a trip from carrier A might be best to collaborate with a trip from carrier B to optimize carrier A's benefits. However, from carrier B's perspective, the same trip might achieve greater profitability if it collaborates with a trip from carrier C.

A simple method to eliminate the conflict options between each carrier is to find the same options between carriers. This is achieved by finding the intersection of the local optimal options across all involved carriers. The solutions that appear in this intersection are considered for the final decision, as they represent the options that all carriers agree upon. This "Intersection Selection" aims to avoid conflict among carriers involved. While it may not yield a globally optimal solution, it ensures a higher degree of mutual satisfaction and feasibility for implementation. The final plans after eliminating the conflict for each carrier show the feasible collaborative solutions.

After summing the benefits of final plan up by the carrier, a matrix, referred to as the "Carrier Benefits Matrix", displays the benefits received from the feasible collaborative solutions for each carrier after resolving conflicts. It quantifies summing up the cost reduction (benefits) from the trip level to the carrier level. This matrix highlights the total benefits realized between carriers, showing the total economic gain from the collaboration for the participants.

Since this matrix shows the economic gains achieved through collaboration, it effectively serves as an index of collaboration effectiveness. Consequently, the 'Carrier Benefits Matrix' can also be named the 'Carrier Collaboration Index Matrix', as it provides a clear indication of the level of successful financial collaboration between the carriers.

IV. Data overview

This section provides a comprehensive overview of the data utilized in this research. The first section, Data Description, introduces the dataset used and highlights the key attributes relevant to the study. In the second section, Data Exploration, various analyses are conducted to discover insights from the pre-processed data. This includes visualizing tours, examining the distribution of trip distances, analyzing the load rate distribution, and presenting an example of the same trip pairs for two carriers. These explorations aim to provide a deeper understanding of the data and its implications for the research.

A. Data description

The dataset utilized in this research comprises extensive freight transportation data, capturing the operational trajectories of hundreds of carriers engaged in pickup and delivery services. The synthesized data was generated by MASS-GT based on freight tour data collection by Central Bureau of Statistics Netherlands (CBS) (de Bok & Tavasszy, 2018; Mohammed et al., 2023). However, the customer data for each carrier, including detailed origin and destination information for cargo, is not available. Only the origin-destination (OD) data are provided and production-consumption (PC) data are not available. Each row within the dataset corresponds to a specific trip undertaken by a carrier, with detailed information spanning 17 attributes.

Key among these attributes are:

- *CARRIER_ID*: This field assigns a unique identifier to each carrier. Within the dataset, a single carrier may undertake multiple tours. A total of 316 carriers are included.
- *TOUR_ID*: This identifier denotes specific tours undertaken by a carrier. For instance, the notation "0_0" in the data frame indicates that the corresponding row refers to the first tour (Tour 0) of the first carrier (Carrier 0).
- *TRIP_ID*: Similar to the TOUR_ID, this identifier marks individual trips within a tour. An entry marked as "0_0_0" signifies the first trip (Trip 0) of the first tour (Tour 0) of the first carrier (Carrier 0).
- *ORIG & DEST*: These identifiers mark the origin and destination of each trip.
- *X_ORIG, X_DEST, Y_ORIG, Y_DEST*: These are coordinates of origin and destination.
- *VEHTYPE*: This refers to the types of vehicles, and different vehicles have corresponding different capacities. A total of seven types of vehicles are included in the dataset.
- *N_SHIP*: This refers to the number of commodities that are inside the truck while travelling between this origin and destination pair.
- *DC_ID*: This identifier shows the position where the carrier's depot belongs to.
- *TOUR_WEIGHT*: The total weight being loaded in the vehicle along the tour.
- *TRIP_WEIGHT*: The weight being transported in the vehicle in the corresponding trip.
- *TRIP_DEPTime*: The trip departure time at the origin.
- *TRIP_ARRTIME*: The trip arrival time at the destination.
- *LOG_SEG*: This refers to the logistic segment, indicating the types of cargo. Nine different types of cargo are listed.

These attributes facilitate a detailed view of freight transport operations at the trip level across various carriers. Additionally, the dataset enables the computation of other operational attributes such as load rate and pickup demand. After filtering out the missing and error data, this research utilizes a preprocessed dataset comprising 95,737 records across 315 carriers.

In the dataset, the "LOG_SEG" attribute categorizes the types of cargo transported along the trips, listing a total of nine different cargo types. The concept of compatible cargo type is crucial, as it determines whether different types of cargo can be safely stored within the same vehicle. To facilitate this, a compatibility matrix, shown in Table II, is assumed and utilized to identify which cargo types can be transported together in the same vehicle. This matrix serves as a guideline for ensuring safe cargo transportation by preventing incompatible cargo combinations. Value one means available and value zero means unacceptable. The matrix is assumed and defined by common sense. For instance, combining food with dangerous materials in the same vehicle is not acceptable. Trip pairs that do not meet the cargo compatibility criteria are excluded from collaborative consideration.

The straightforward method to avoid complex calculations involves using a predefined cost per kilometre for the seven different vehicle types included in MassGT (de Bok & Tavasszy, 2018; Mohammed et al., 2023). The unit of this cost is expressed in €/km. All parameters are summarized in Table III.

B. Data exploration

In this section, a journey is undertaken to discover insights from the pre-processed dataset described in the previous section.

Each row of the data is based on the trip level, the distance for each trip can also be explored. The histogram can distribute the data into bins, identifying the variability of trip distances within the data. The determination of bin size would influence the visualization to reflect the trip distance distribution. For this research, the bin size was determined using Sturges' rule. Sturges' rule suggests setting the number of bins to $1 + \log_2(N)$, where N is the number of data points which is 95,737. By employing this approach, the bin size is set to 18. Figure 3 shows the trip distance distribution of the data. The figure indicates frequent trips within the initial bins, particularly in the 0-25,121 meter range. The first bin include for more than 23,000 trips, suggests that most of the journeys are relatively short. As distances increase, a marked decline in the number of trips is observed. Extremely long journeys, such as those about 400,000 meters are infrequent. This could reflect specific logistic requirements for long distances within the data.

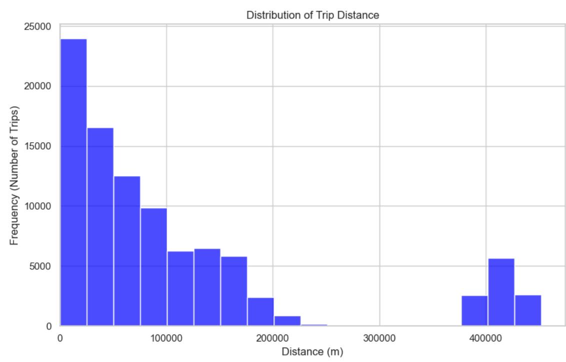


Fig. 3. Distribution of trip distance

The trip load rate is an indicator of the efficiency of freight transportation. A low load rate for carriers indicates inefficiency and potential capacity for sharing with partners in horizontal collaboration.

TABLE II
COMPATIBLE CARGO TYPE MATRIX

	Food (general cargo)	Miscellaneous (general cargo)	Temperature controlled	Facility logistics	Construction logistics	Waste	Parcel (consolidated flows)	Dangerous	Parcel (deliveries)
Food (general cargo)	1	1	1	0	0	0	1	0	1
Miscellaneous (general cargo)	1	1	1	0	0	0	1	0	1
Temperature controlled	1	1	1	0	0	0	1	0	1
Facility logistics	0	0	0	1	1	0	0	0	0
Construction logistics	0	0	0	1	1	0	0	0	0
Waste	0	0	0	0	0	1	0	0	0
Parcel (consolidated flows)	1	1	1	0	0	0	1	0	1
Dangerous	0	0	0	0	0	0	0	1	0
Parcel (deliveries)	1	1	1	0	0	0	1	0	1

TABLE III
COST PER KILOMETRE FOR DIFFERENT VEHICLE TYPES

VEHTYPE	CostPerKm
Truck_Small	0.1954
Truck_Medium	0.3404
Truck_Large	0.5724
TruckTrailer_Small	0.2244
TruckTrailer_Large	0.5144
TractorTrailer	0.4274
SpecialVehicle	0.4274

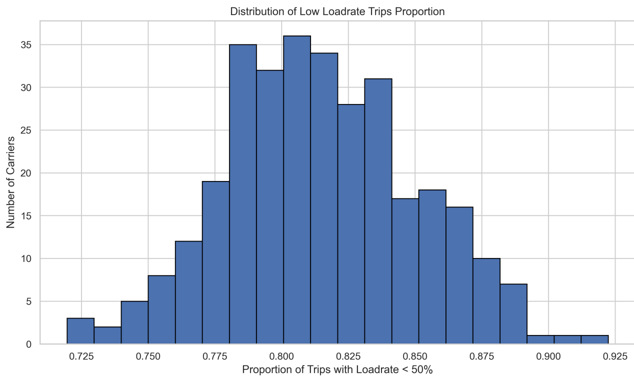


Fig. 4. Distribution of the proportion of trips with load rate below 50% among carriers

Figure 4 illustrates the distribution of the proportion of trips where the load rate is below 50% for various carriers. The horizontal axis represents the proportion of such trips, while the vertical axis shows the number of carriers. The chart highlights the potential for collaboration to optimize load utilization.

The figure reveals significant variability in the proportion of low-load-rate trip journeys, with all carriers within the data exhibiting low-load trips above 70%. Most carriers having a low load rate trips proportion between 0.775 and 0.85. This indicates that for the majority of carriers, about 77.5% to 85% of their trips have a load rate below 50%. This high proportion of low load rate trips highlights the current low utilization of available capacity and suggests significant potential for optimization and collaboration.

If reallocation is not considered, the same origin-destination (OD) trips between different carriers reveal a close relationship and potential collaborative opportunities, as introduced in Section III. Figure 5 illustrates the frequency of trips for different carriers between origin and destination pairs. Each line represents a trip path taken by multiple carriers. Each subplot represents a different range of OD counts: 1 to 50, 51 to 100, 101 to 200, and 201 and above. The colour gradient, ranging from light blue to dark blue, represents the frequency of different carriers recording the same origin-destination (OD) trip. Darker shades indicate higher trip frequencies, highlighting routes with more frequent travel by different carriers. This visualization shows the potential for collaboration on these identical high-frequency OD trips.

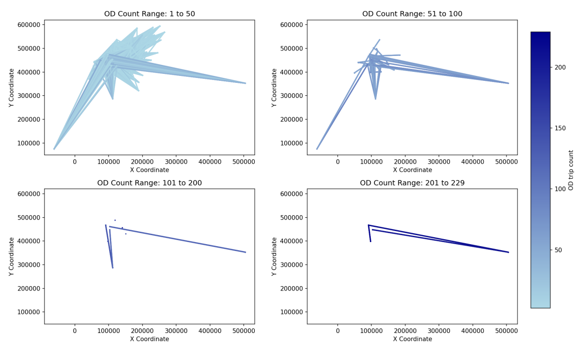


Fig. 5. Frequency of same OD trips by different carriers

The visualization presented in Figure 5 offers a comprehensive view of the frequency of the same OD trips. Here an simple same OD trip example are given, Figure 6 focuses on the trip level, illustrating two tours by carrier 65 and carrier 232. These tours follow the same trip route from node 268 to node 3840. Visiting the same nodes is an insight into horizontal collaboration, such as sharing capacity and enhancing transportation efficiency. For instance, one carrier could handle the transport task for a shared trip, allowing the other carrier to optimize operations and improve efficiency. This example highlights carriers visiting the same OD trip, but generally, trip pairs within acceptable proximity also offer potential for collaboration. Further details on determining collaboration opportunities will be explored in the next section.

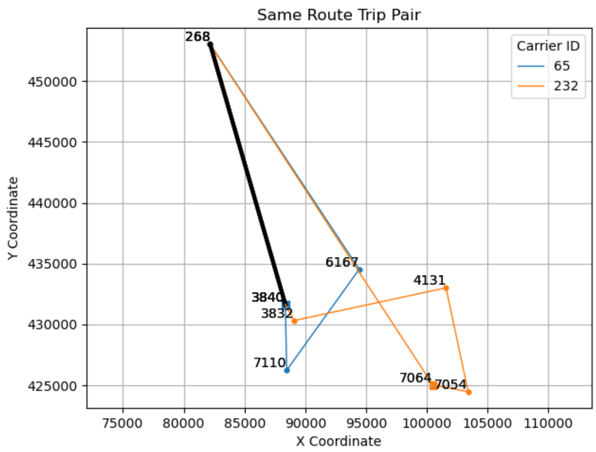


Fig. 6. Visualization trip pair routes from carrier 65 and carrier 232

V. Implementation

This section constructs several scenarios based on the parameters and variables included in the methodology to facilitate a deeper examination of the results. These scenarios serve as frameworks for testing various understanding of the implications of the collaboration index from different perspectives. By applying these scenarios, the result analysis can discover insights into the collaboration index.

The limit range for allocation, discussed in Section III-A2, allows for routing adjustments between trips. This predefined acceptable range for allocation influences the number of collaboration opportunities available, which in turn affects the collaboration index between carriers. By adjusting this range, the impact on the collaboration index is investigated, providing insights into how flexible routing parameters can enhance or limit the potential for effective collaboration between carriers.

The heuristic local optimization provides a simple approach to aggregate trip-level collaboration modes into the carrier-level index. After the collaboration options are sorted according to the five metrics presented in Section III-B1, two additional variables are considered before the selection procedure: random selection rate and cost reduction threshold.

The random selection rate is employed to illustrate that carriers might not consider all available options when exploring collaboration. Instead, there is a probability that they will choose to collaborate at a specific rate. Considering this rate of selection introduces additional randomness into the computation of benefits. Moreover, it helps in demonstrating how the benefits fluctuate based on variations in the selection rate.

The threshold is employed to filter out less impact collaboration options, allowing carriers to concentrate their resources and efforts on collaborations that offer significant cost savings. This cost reduction threshold is used to assess the viability of potential collaborative options. The cost reduction rate is defined as the percentage of cost savings relative to the total cost of the impacted routes. Specifically, a collaboration option is only considered viable if the cost savings achieved surpass a predefined percentage threshold.

A total of 54 scenarios were generated through three variables: limit range, random selection rate, and cost reduction thresholds. The unique combinations derived from varying levels of each variable—limit range (0, 500, 1000 km), random selection rate (100%, 80%, 60%), and cost reduction thresholds (0%, 10%, 20%, 30%, 40%, 50%). The exploration of these scenarios helps investigate how each variable impacts the performance and cost-saving potential of collaboration. This analysis also contributes to the development of the collaboration index. The detailed analysis is presented in Section VI.

The ’Carrier Collaboration Index Matrix’ is the output of this research, providing a direct relationship between carriers, indicating the potential collaboration opportunities and the level of benefits that may shared with the participants.

The ’Carrier Collaboration Index Matrix’ is the key output of this research. It provides a direct relationship between carriers, indicating potential collaboration opportunities and the level of benefits that would be shared among the participants. This matrix serves as a tool for carriers to identify and evaluate possible partnerships, enhancing their decisions regarding collaborative strategies and optimizing overall operational efficiency. The detailed application for the collaboration index is introduced in Section VI-B.

VI. Result Analysis

In this section, the scenarios developed in Section V are applied. In the first section, the results, including the total benefits and cost-saving rate under different scenarios, are analyzed through sensitivity analysis. This analysis provides insights into the effectiveness and robustness of the proposed strategies under various conditions.

In the second section, the collaboration index obtained from this research is presented. Additionally, several analyses of the index are conducted, offering relevant advice on horizontal collaboration strategies based on the findings. And an in-depth analysis based on the collaboration index for interpretation is followed. This comprehensive evaluation aims to enhance the understanding and application of the collaboration index in real-world transportation logistics.

A. Sensitivity analysis

This section presents the effect of limit range, random selection rate and cost reduction threshold on the performance of total benefits of the final collaboration plan and total cost saving rate. All of them are based on the carrier level.

The total benefits of the final collaboration plan are excluded the option conflict as introduced in Section III-B1. The value of total benefits shows the total amount of benefits could gained from the collaboration for the data under the specific scenarios. The total cost saving rate is calculated by the following formula.

$$\text{Total cost saving rate } (\%) = \left(\frac{\text{Total\_Cost\_saving}}{\text{Total\_Cost}} \right) \times 100 \tag{1}$$

The $Total\_Cost$ refer to the total cost for all the trips involved in the data without collaboration. The total cost saving rate show the percentage of benefits account for the total cost, which emphasize the effectiveness of collaboration.

Table IV shows the result of total benefits and cost savings varies different limit range (0m, 500m, 1000m), random selection rate (100%, 80%, 60%) and cost reduction threshold (0%, 10%, 20%, 30%, 40%, 50%).

1) Effect of the Limit Range: The effect of the Limit Range on the total cost-saving rate is illustrated in Figure 7. Each subplot corresponds to a different Random Selection Rate (100%, 80%, 60%), and the lines represent different cost reduction thresholds. It is evident from the figure that as the Limit Range increases from 0 to 1000, the total cost-saving rate generally shows an upward trend. This suggests that increasing the limit range expands the decision-making space, allowing for more reallocation considerations and including more potential trips in the determination process. Consequently, this enhances benefits and cost savings, as a larger limit range facilitates the identification and optimization of more collaborative opportunities.

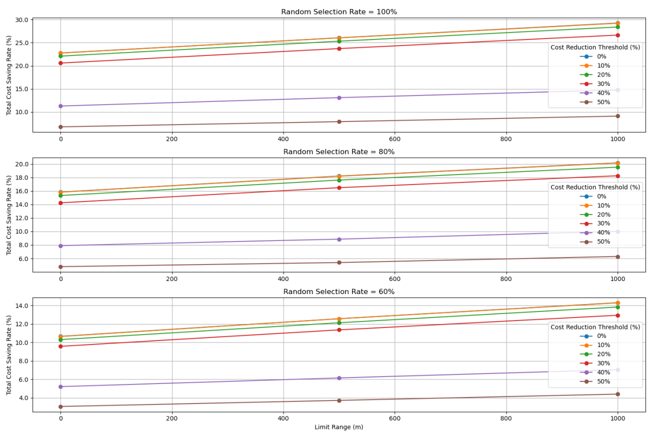


Fig. 7. The relationship between Total Cost Saving Rate and Limit Range across different Random Selection Rate and Cost Reduction Thresholds

2) Effect of the Random Selection Rate: Figure 8 illustrates the total cost saving rate as a function of the random selection rate (%) for various cost reduction thresholds. Each subplot corresponds to a different Limit Range (0m, 500m, 1000m),

TABLE IV
THE TOTAL BENEFITS AND COST SAVINGS UNDER DIFFERENT LIMIT RANGE (0M, 500M, 1000M) RANDOM SELECTION RATE (100%, 80%, 60%) AND COST REDUCTION THRESHOLD (0%, 10%, 20%, 30%, 40%, 50%)

Limit Range (m)	Random Selection Rate (%)	Cost Reduction Threshold (%)	Total Cost Saving Rate (exclude option conflict) (%)
0	100	0	22.77
0	100	10	22.74
0	100	20	22.09
0	100	30	20.59
0	100	40	11.29
0	100	50	6.78
0	80	0	15.84
0	80	10	15.82
0	80	20	15.33
0	80	30	14.25
0	80	40	7.90
0	80	50	4.78
0	60	0	10.65
0	60	10	10.63
0	60	20	10.30
0	60	30	9.57
0	60	40	5.21
0	60	50	3.06
500	100	0	26.06
500	100	10	26.03
500	100	20	25.32
500	100	30	23.72
500	100	40	13.10
500	100	50	7.90
500	80	0	18.22
500	80	10	18.19
500	80	20	17.62
500	80	30	16.49
500	80	40	8.86
500	80	50	5.39
500	60	0	12.56
500	60	10	12.55
500	60	20	12.12
500	60	30	11.35
500	60	40	6.15
500	60	50	3.72
1000	100	0	29.25
1000	100	10	29.20
1000	100	20	28.39
1000	100	30	26.64
1000	100	40	14.75
1000	100	50	9.10
1000	80	0	20.17
1000	80	10	20.14
1000	80	20	19.53
1000	80	30	18.26
1000	80	40	10.02
1000	80	50	6.29
1000	60	0	14.29
1000	60	10	14.27
1000	60	20	13.82
1000	60	30	12.93
1000	60	40	7.04
1000	60	50	4.40

showcasing how cost savings fluctuate across different selection rates and thresholds.

The Random Selection Rate is used for showing the carriers may not consider all collaboration options for collaborate. The randomness add to the model also help to indicate the fluctuate of benefits. When the Random Selection Rate decreases (from 100% to 60%), irrespective of the Limit Range, the total cost saving rate tend to decrease. This indicates that a higher random selection rate improves project benefits and cost efficiency, because greater randomness offers more potential favorable collaboration combinations.

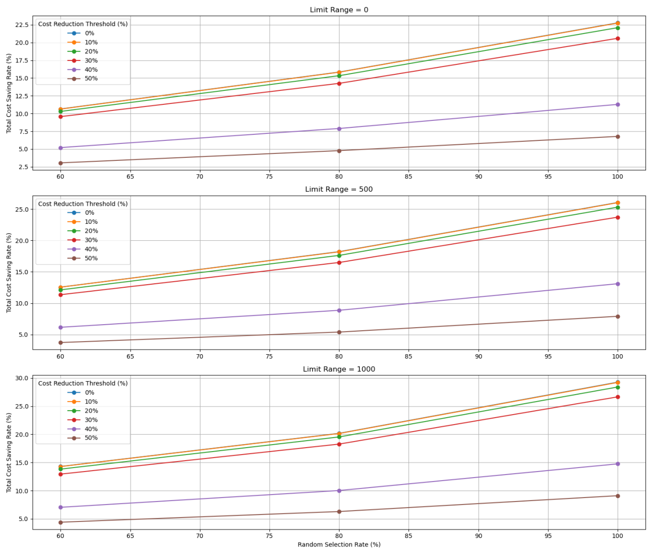


Fig. 8. The relationship between Total Cost Saving Rate and Random Selection Rate across different Limit Ranges and Cost Reduction Thresholds

3) Effect of the Cost Reduction Threshold: The Cost Reduction Threshold is employed to enable carriers to concentrate their resources on collaborations that yield high cost savings. However, since the cost calculations in this research consider only distance and do not account for other realistic expenses such as wages and storage costs, collaborations with low cost savings might result in negative revenue in actual practice. Implementing this threshold helps to mitigate such risks by ensuring that only those collaborations with significant cost-saving potential are pursued, thereby maximizing the likelihood of positive financial outcomes for the carriers.

Figure 9 shows the effect of the Cost Reduction Threshold on the total cost-saving rate. As the Cost Reduction Threshold increases, both the total benefits and the total cost-saving rate generally decline. This is because higher Cost Reduction Thresholds eliminate collaboration options with lower benefits, thus reducing the overall benefits. The slope of the decrease continues to steepen, with a significant drop observed when the Cost Reduction Threshold exceeds 30%. This indicates that beyond this threshold, the exclusion of potential collaborations with relative low cost savings markedly impacts the total cost-saving rate, highlighting the trade-off between stricter criteria and overall savings.

For example, when the Limit Range is set at 1000 meters and the Random Selection Rate at 100%, the slopes are recorded as (-0.005, -0.081, -0.174, -1.190, -0.564). These values represent the slopes, and their progression illustrates a gradual transition from a small negative value to increasingly larger negative values. The trend in these slopes indicates an initially slow decline that accelerates rapidly before slightly decelerating, yet it still maintains a rapid downward trajectory. The significant drop observed between the 30% and 40% thresholds may indicate that many of the collaboration decisions occur within this range for the dataset explored in this research. One contributing factor to the observed decline is that within the 30 to 40 Cost Reduction Threshold, many collaboration plans involving one of the carriers only include two nodes throughout the entire tour. Specifically, the entire transport task of the tour could be managed by the partner carrier if collaboration occurs. Although reallocation led to increased costs due to detours, the collaboration plan still yielded positive benefits. This arrangement facilitated the sharing of reduced costs generated through collaboration, thereby enabling profit sharing between the carriers.

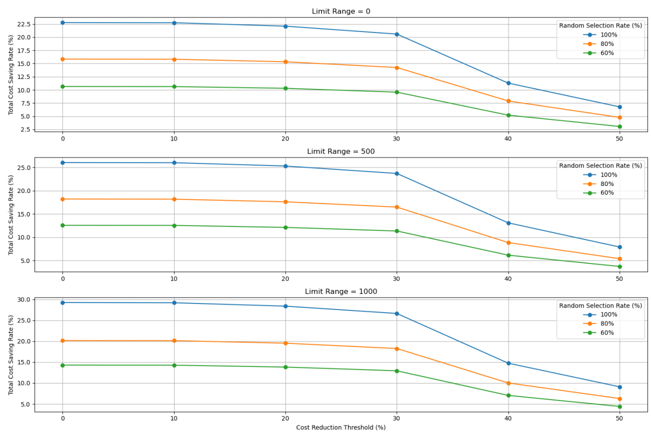


Fig. 9. The relationship between Total Cost Saving Rate and Cost Reduction Thresholds across different Limit Ranges and Random Selection Rate

B. Collaboration index analysis

Section VI-A investigated the effects of three variables on the benefits and cost-saving rate of the collaboration. Based on these findings, a Limit Range of 1000m, a Random Selection Rate of 100%, and a Cost Reduction Threshold of 20% are applied in this section. These values allow for reallocation between trips, with carriers following all collaboration decision suggestions and focusing on high cost-saving collaboration options by following a reasonable threshold. The output of the collaboration index, derived from these parameters, will be analyzed in this section, providing insights into the horizontal collaboration strategies and the suggestions provided to the carriers.

The data for this research includes a total of 316 carriers. The resulting "Carrier Collaboration Index Matrix" has dimensions of 316 by 316. A part of the matrix is shown in Table V. This example illustrates how the collaboration index can be utilized in freight transportation and highlights the suggestions derived from the collaboration index. By examining this matrix, carriers can identify their optimal partners for collaboration, focusing on those with high potential benefits. This enables carriers to make decisions and implement collaboration solutions that increase efficiency and save costs.

TABLE V
COLLABORATION INDEX FOR CARRIERS WITH IDS RANGING FROM 20 TO 30

Carrier_ID	20	21	22	23	24	25	26	27	28	29	30
20	0.00	375.06	245.42	208.17	488.05	0.00	0.00	0.00	0.00	135.50	0.00
21	375.06	0.00	275.32	1277.09	360.55	0.00	0.00	0.00	0.00	0.00	86.74
22	245.42	275.32	0.00	994.52	55.45	0.00	0.00	0.00	0.00	114.41	42.59
23	208.17	1277.09	994.52	0.00	2104.28	0.00	0.00	0.00	0.00	202.54	169.74
24	488.05	360.55	55.45	2104.28	0.00	0.00	0.00	0.00	0.00	722.29	222.44
25	0.00	0.00	0.00	0.00	0.00	0.00	407.03	461.56	593.79	18.46	0.00
26	0.00	0.00	0.00	0.00	0.00	407.03	0.00	1457.30	607.79	0.00	318.07
27	0.00	0.00	0.00	0.00	0.00	461.56	1457.30	0.00	1634.04	129.72	12.18
28	0.00	0.00	0.00	0.00	0.00	593.79	607.79	1634.04	0.00	433.96	193.20
29	135.50	0.00	114.41	202.54	722.29	18.46	0.00	129.72	433.96	0.00	495.49
30	0.00	86.74	42.59	169.74	222.44	0.00	318.07	12.18	193.20	495.49	0.00

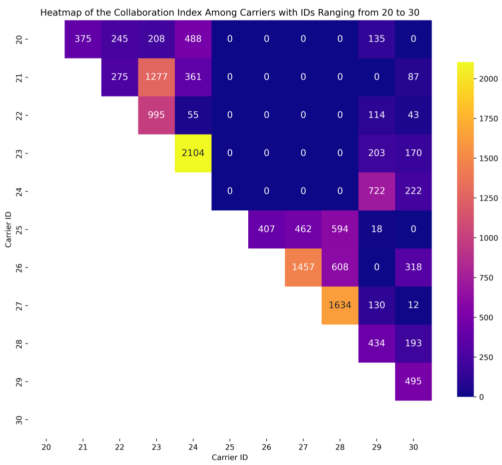


Fig. 10. Heatmap of the collaboration index among carriers with IDs ranging from 20 to 30

The collaboration index is shown in Table V includes carriers with IDs ranging from 20 to 30. For a carrier willing to collaborate, the collaboration index reveals potential opportunities. The row corresponding to the carrier's ID provides this information. For example, if carrier 25 wants to check potential partners for collaboration, the values in row 25 of the Collaboration Index Matrix will be provided. The value of the collaboration index indicates the amount of benefits (€) that can be shared when collaborating with the corresponding carrier ID in the column. A higher value of the collaboration index signifies a greater potential for profit sharing in the partnership, guiding carriers towards the most beneficial collaborations.

The heat map displayed in Figure 10 illustrates the collaboration index (CI) among carriers, highlighting the varying potential for collaboration between different carriers. The colour scale transitions smoothly from dark purple, representing lower CI values, to bright yellow, indicating higher CI values. Brighter colours denote the higher potential for collaboration and greater benefits shared following its application.

On behalf of carrier 25, the potential shared benefits with carriers 26, 27, and 28 are substantial (407, 462, 594), while the benefits generated from collaboration with carrier 29 are relatively low (18). Although the Cost Reduction Threshold is set at 20%, which eliminates options with slight benefits from the solutions, carrier 25 still needs to assess whether collaboration with carrier 29 would be efficient enough to bring positive benefits in practice. Additionally, detailed decisions for collaboration with different partners are accessible to the carrier through the Collaborative Decision Support System. This system provides comprehensive information on trip-level details, modes of collaboration, collaborative planned routes, and the loading or unloading cargo solutions for vehicles, enabling carrier 25 to make well-informed decisions regarding potential partnerships.

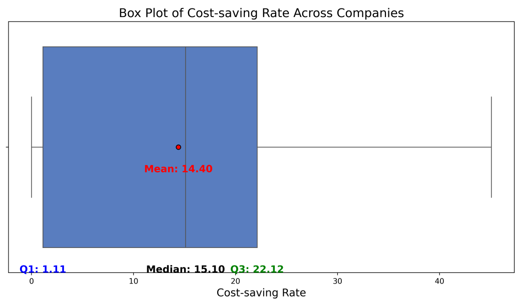


Fig. 11. Distribution of cost-saving rates across companies from horizontal collaboration

The collaboration index indicates the amount of benefits that can be shared between partners when they collaborate. The cost-saving rate for each participant in the collaboration is also influenced by the total cost across the entire tours in the data. The cost-saving rate helps participants better understand the expected cost reduction impact of horizontal collaboration if they collaborate based on the collaboration plan suggestions.

The box plot shown in Figure 11 illustrates the distribution of the cost-saving rate across all companies engaged in horizontal collaboration. The first quartile (Q1), which indicates that the lowest 25% of data points fall below this value, is 1.11%. The median (Q2) is 15.10%, providing a central point that separates the lower and upper halves of the data. The mean cost-saving rate is 14.40%, closely aligning with the median, suggesting a relatively symmetrical distribution of savings. The third quartile (Q3), where the lowest 75% of data points fall below this value, is 22.12%. These quartiles highlight the variability in cost-saving rates among companies. On average, participants in the horizontal collaboration benefit from a 14.4% cost-saving rate, demonstrating the potential advantages of such partnerships.

1) Collaboration index interpretation: While this research develops a two-step method for examining and identifying horizontal collaboration using the collaboration index, an in-depth analysis of the collaboration index involves explaining and interpreting how various attributes influence its value. In this section, the impact of geographical location and vehicle types is investigated.

In Section III-A2, geographical location is used to determine the distance between two trips and whether the origin and destination nodes are within a certain range. Here, a higher-level analysis focuses on the overall node locations, investigating the similarity in node locations between the two carriers. This analysis helps to understand how geographical proximity contribute to potential collaboration opportunities.

Initially, the entire spatial area in the dataset is divided into grids of a predetermined size, referred to as the "grid size." Each grid represents a distinct two-dimensional spatial region. This process facilitates the spatial partitioning necessary to analyze the distribution of nodes among carriers.

For each carrier, the number of nodes within each grid is counted, resulting in a distribution matrix for each carrier that reflects spatial distribution of nodes within the defined grids. This step transforms spatial data into a quantifiable format suitable for further analysis.

After that, to explore the similarity between the two carriers, cosine similarity is employed due to its simplicity in comparing distribution similarities and its efficiency in short-time computations (Xia et al., 2015). The formula for cosine similarity is as follows. x and x' refer to the two vectors.

$$\text{Cosine Similarity} = \frac{\sum_{i=1}^{d} x_i \times x'_i}{\sqrt{\sum_{i=1}^{d} x_i^2} \times \sqrt{\sum_{i=1}^{d} x_i'^2}} \quad (2)$$

This formula measures the directional similarity between two vectors, regardless of their magnitude. By focusing on direction rather than length, cosine similarity emphasizes proportional relationships between node distributions. Before computation, each distribution matrix is flattened into a one-dimensional vector. The relationship between cosine similarity and the collaboration index is illustrated in Figure 12.

The cosine similarity values are computed for all carrier pairs to investigate the correlation with their collaboration index. The Pearson correlation is applied to measure the linear relationship between cosine similarity and the collaboration index (CI), indicating the relationship between the geographical location of nodes and the collaboration index.

Several different grid size are given, and all the correlation coefficient are shown in Table VI. For the grid size 500, 1000, 2000 and 3000, the result of coefficient are about 0.50,

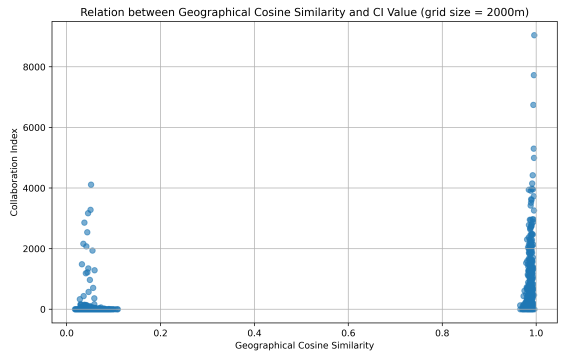


Fig. 12. Relationship between geographical cosine similarity and collaboration index across carrier pairs (grid size = 2000m)

indicating a moderate positive relationship between cosine similarity and the CI value. A positive correlation coefficient suggests that as cosine similarity increases, the CI value tends to increase as well. This implies that carriers with similar spatial distributions may also have higher collaboration potential. All of the coefficient falls within the range of 0.3 to 0.7, which is typically considered a moderate positive correlation. While this is not a very strong correlation, it does indicate a noticeable connection between the two variables.

TABLE VI
CORRELATION COEFFICIENTS BETWEEN GEOGRAPHICAL COSINE SIMILARITY AND COLLABORATION INDEX AT VARIOUS GRID SIZES

Grid size (m)	500	1000	2000	3000	5000
Correlation coefficient	0.57	0.54	0.51	0.49	0.39

Examining vehicle types allows for an assessment of how vehicle characteristics influence the collaboration index. This analysis explores whether the types of vehicles owned by carriers in the suggesting collaboration plan are related to the collaboration index.

Firstly, the same vehicle rate between carriers is computed to show the percentage of same vehicles in the collaboration plans between specific carrier pairs. A higher same vehicle rate indicates a greater number of collaborations suggested between trips operated by the same types of vehicles. Figure 13 shows the result. The Pearson correlation coefficient is also applied here to examine the relationship between the same vehicle rate and the collaboration index. The result of the Pearson correlation coefficient is -0.028, suggesting a very weak negative linear relationship between the two variables. The value being close to zero implies that there is almost no linear correlation, which means the same vehicle rate are not linearly related with the CI.

VII. DISCUSSION

This research focused on the area of partner selection in road freight transport horizontal collaboration, addressing a

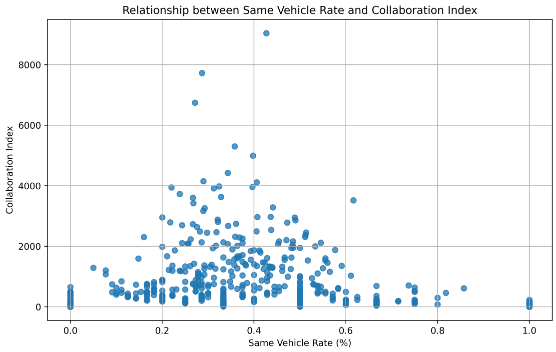


Fig. 13. Relationship between same types of vehicle rate and collaboration index across carrier pairs

notable gap in the current literature. Previous work has often neglected the operational domain, specifically the selection of partners based on freight transportation data. While many studies employ complex optimization approaches to determine collaboration plans between two participants, there has been less emphasis on identifying potential collaborations from a broader perspective.

In addressing the research questions outlined in Section I-D, each question is resolved as follows.

For the sub-question 1, the "collaboration index" offers significant advantages to various stakeholders. For carriers participating in collaboration, the index facilitates the efficient selection of partners with high expected benefits. The Collaborative Decision Support System further provides detailed collaborative plans for each carrier, including vehicle routes, modes of collaboration, and loading or unloading operations for vehicles. This ensures that carriers can optimize their resources and achieve cost savings through strategic collaborations.

For governments and authorities, the collaboration index serves as a valuable reference for assessing the extent of collaboration within the freight transportation sector. It provides insights into current collaboration levels and identifies areas with potential for increased cooperative efforts, thereby informing policy and regulatory decisions to support and enhance freight transport efficiency.

Coordinators, who act as neutral third parties in the Collaborative Transport Networks, can use the collaboration index to manage and promote collaboration opportunities among carriers. By sharing these opportunities based on the collaboration index, coordinators help participants identify the most beneficial partnerships. Additionally, the coordinators ensure the confidentiality of sensitive company information, maintaining trust and neutrality in the collaboration process.

Sub-question 2 would answer from the Section III-A and Section III-B. The collaboration determination and the collaboration index formulation involve consideration of various attributes and factors. In the collaboration determination part, several attributes are considered: the distance between trips, the types of cargo, depot locations, vehicle loads, loading and unloading operations, and vehicle capacities. The formulation of the collaboration index incorporates additional variables to enhance the precision and applicability of the index. These include the total profit derived from collaboration, the frequency of collaboration, bias in departure times, limit ranges, the rate of random selection, and the thresholds for cost reduction. All these attributes and factors be considered in the collaboration index research to reflect the potential for collaborative logistics.

To address sub-question 3, which seeks to quantitatively identify potential collaboration opportunities among carriers, the collaboration index was formulated. This index was developed through processes of collaboration determination and collaboration index formulation. The collaboration determination process utilized a decision tree to identify four modes of collaboration at the trip level. The collaboration index formulation aggregated these trip-level collaboration modes into a carrier-level quantitative evaluation. Among it, the heuristic local optimization approach was applied to solve the two types of conflicts behind the options for each carrier. As a result, the local optimized collaboration solutions were obtained and the collaboration index quantitatively demonstrated the potential benefits of these feasible collaborative solutions.

The key contribution of this research proposes an innovative collaboration index for carriers, identifying horizontal collaboration opportunities in road freight transport. The collaboration decision tree is employed to determine four modes of collaboration at the trip level from freight transport data. Subsequently, a heuristic local optimization approach aggregates these modes at the carrier level, formulating a quantitative collaboration index.

The findings from the result include the sensitivity analysis and further in-depth analysis of the collaboration index. In Section VI-A three variables were examined for the collaboration index: limit range, random selection rate, and cost reduction threshold. Sensitivity analysis was conducted to assess their effects on the total cost-saving rate. The results indicated that an increase in the limit range could generate more opportunities for collaboration, with a corresponding upward trend in total cost savings. A decrease in the random selection rate resulted in a reduction in the total cost-saving rate. Conversely, as the cost reduction threshold increased, the total cost-saving rate generally declined, with a significant drop observed when the threshold exceeded 30%. This analysis enhances the understanding of the collaboration index by applying it across various scenarios with the three variables.

In this dataset, companies participating in the horizontal collaboration can achieve an overall average cost-saving rate of 14.4%. The lowest 25% of companies experience a cost-saving rate of 1.11%, while the highest 25% achieve a cost-saving rate of 22.12%. Further research indicates a moderately positive correlation between the geographical similarity of node locations and the collaboration index among carrier pairs. However, the use of same vehicle types between carrier pairs does not exhibit a linear relationship with the collaboration index.

There are still some shortages and limitations in this research. Due to the lack of detailed customer information, specifically the trade level production-consumption data, this study only considered weight variations between node visits and did not account for changes in cargo quantity or volume. The research assumed that the cargo carried in a vehicle during its entire tour is homogeneous. Also, although the compatible cargo types were assumed in this research, the application of the compatibility matrix still needs to be examined and investigated through real-world operations.

Additionally, the development of the collaboration index was based on a heuristic local optimization approach, guided by predefined selection metrics. While this method proves efficient, it falls short of achieving global optimization. It does not minimize the total costs incurred by all participants. As a result, the final collaborative solutions for each carrier do not represent globally optimized outcomes. Moreover, the distribution of profits among participants has not been addressed in this research.

Future studies should aim to utilize operational data alongside customer data, including detailed cargo flows, and account for the volume and quantity of cargo to identify horizontal collaboration opportunities. This comprehensive approach would yield more practical and applicable results for modes of collaboration.

The four collaboration modes identified through the decision tree introduced in Section III-A2 not only categorize the collaboration types but also assign labels on trip-level. These labels can be used as target values for machine learning approaches, allowing for further development with additional preferences.

Additionally, this research employed a relatively simple decision-making process to optimize collaborative solutions. Future research could compare this straightforward approach with other optimization models or data-driven methods to evaluate their effectiveness and efficiency. Such comparisons could provide valuable insights into the advantages and limitations of various methodologies in optimizing collaborative freight transport.

The collaboration index introduced in this study also presents an interesting area for further exploration. Future work could investigate its real-world applications, for example, the profit distribution based on the contribution, examining how the collaboration index be integrated with existing transportation networks. Exploring these could significantly enhance the practical and theoretical understanding of the collaboration index in freight transportation.

VIII. Conclusion

This research aimed to identify potential horizontal collaboration opportunities with the defined collaboration index. By focusing on partner selection through the analysis of freight transportation data, this study fills a significant gap in the existing literature. The collaboration index was developed to quantitatively assess potential collaboration opportunities among carriers. It utilizes a collaboration decision tree and heuristic local optimization to aggregate data from the trip-level to the carrier-level evaluation. Several attributes were considered, including trip route, depot position, load weight, operational tasks, vehicle types, distance, and cost. The integration of these attributes facilitated the construction of the comprehensive approach. Different scenarios defined by three variables limit range, random selection rate, and cost reduction threshold, influence the cost savings achieved through collaboration. The sensitivity analysis highlights the impact of these variables on cost savings.

According to the research data, if the carriers engage in horizontal collaboration, they have the potential to achieve an average cost-saving of 14.4%, with the top 25% of carriers realizing savings of up to 22.12%. The similarity in geographical node locations between carriers is associated with the value of the collaboration index. However, the same types of vehicles do not appear to have a significant relationship with the collaboration index.

The implications of this research are multifaceted. For carriers, the collaboration index serves as a strategic tool to select partners, potentially leading to optimized operations and cost reductions. The collaboration index has become a fundamental quantitative value that illustrates the potential benefits and cost reductions, serving as a motivation for participants to engage in collaboration. Governments and authorities can leverage the index as a benchmark for the extent of collaboration within road freight transportation, which could guide policy decisions aimed at enhancing transport efficiency. Furthermore, coordinators within Collaborative Transport Networks can utilize the index to suggest and manage beneficial partnerships among carriers, ensuring the neutrality and confidentiality required for successful collaboration.

The research acknowledges its limitations, such as the lack of detailed trade-level data and the reliance on a heuristic local optimization approach that does not achieve global cost minimization. Future research should, therefore, expand to include production-consumption data to investigate the collaboration index further. Comparing the current model with other optimization methods could also provide deeper insights into the collaborative strategies in freight transport.

While the collaboration index developed in this research serves as a quantitative tool for identifying horizontal collaboration in road freight transport, the collaboration index lead to a practical impact on the quantitative identification of collaboration partners, seeking to find partners to receive win-win. The adaption of collaboration index in real-world application such as collaboration decision support system (CDSS) have strong potential and the research of collaboration index could be further improve.

The collaboration index developed in this research provides a quantitative framework for identifying potential partners in horizontal collaboration within road freight transport. This index facilitates the strategic pairing of companies aiming to achieve mutually beneficial. When integrated into a Collaboration Decision Support System (CDSS), the collaboration index demonstrates significant potential to enhance real-world applications. Further development and research into the collaboration index would essential to maximize its efficiency and broaden its applicability in road freight logistics. Such advancements would expected to pave the way for a more efficient road freight transport with horizontal collaboration.

References

Agarwal, R., & Ergun, Ö. (2010). Network Design and Allocation Mechanisms for Carrier Alliances in Liner Shipping [Publisher: INFORMS]. *Operations Research*, *58*(6), 1726–1742. https://doi.org/10.1287/opre.1100.0848

Amer, L. E., & Eltawil, A. B. (2015). Analysis of quantitative models of horizontal collaboration in supply chain network design: Towards "green collaborative" strategies. *2015 International Conference on Industrial Engineering and Operations Management (IEOM)*, 1–10. https://doi.org/10.1109/IEOM.2015.7093759

Audy, J.-F., D'Amours, S., & Rönnqvist, M. (2007). Business Models for Collaborative Planning in Transportation: An Application to Wood Products. In L. M. Camarinha-Matos, H. Afsarmanesh, P. Novais, & C. Analide (Eds.), *Establishing the Foundation of Collaborative Networks* (pp. 667–676). Springer US. https://doi.org/10.1007/978-0-387-73798-0_72

Audy, J.-F., Lehoux, N., D'Amours, S., & Rönnqvist, M. (2012). A framework for an efficient implementation of logistics collaborations. *International Transactions in Operational Research*, *19*(5), 633–657. https://doi.org/10.1111/j.1475-3995.2010.00799.x

Awasthi, A., Adetiloye, T., & Crainic, T. G. (2016). Collaboration partner selection for city logistics planning under municipal freight regulations. *Applied Mathematical Modelling*, *40*(1), 510–525. https://doi.org/10.1016/j.apm.2015.04.058

Bahinipati, B. K., Kanda, A., & Deshmukh, S. G. (2009). Horizontal collaboration in semiconductor manufacturing industry supply chain: An evaluation of collaboration intensity index. *Computers & Industrial Engineering*, *57*(3), 880–895. https://doi.org/10.1016/j.cie.2009.03.003

Beliën, J., Boute, R., Creemers, S., Bruecker, P., Gijsbrechts, J., Padilla Tinoco, S. V., & Verheyen, W. (2017). Collaborative Shipping: Logistics in the Sharing Economy.

Cruijssen, F., Dullaert, W., & Fleuren, H. (2007). Horizontal Cooperation in Transport and Logistics: A Literature Review. *Transportation Journal*, *46*. https://doi.org/10.5325/transportationj.46.3.0022

Dahl, S., & Derigs, U. (2011). Cooperative planning in express carrier networks — An empirical study on the effectiveness of a real-time Decision Support System. *Decision Support Systems*, *51*(3), 620–626. https://doi.org/10.1016/j.dss.2011.02.018

de Bok, M., & Tavasszy, L. (2018). An empirical agent-based simulation system for urban goods transport (MASS-GT). *Procedia Computer Science*, *130*, 126–133. https://doi.org/10.1016/j.procs.2018.04.021

Deng, S., Zhou, D., Wu, G., Wang, L., & You, G. (2023). Evolutionary game analysis of three parties in logistics platforms and freight transportation companies' behavioral strategies for horizontal collaboration considering vehicle capacity utilization. *Complex & Intelligent Systems*, *9*(2), 1617–1637. https://doi.org/10.1007/s40747-022-00873-9

Ergun, O., Kuyzu, G., & Savelsbergh, M. (2007). Reducing Truckload Transportation Costs Through Collaboration [Publisher: INFORMS]. *Transportation Science*, *41*(2), 206–221. https://doi.org/10.1287/trsc.1060.0169

Essam, L., & Eltawil, A. (2014). Collaborative Sustainable Supply Chain Network Design: State Of The Art and Solution Framework.

Eurostat. (2023, October). Road freight transport by journey characteristics. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Road_freight_transport_by_journey_characteristics#Empty_runnings_of_road_freight_vehicles

Feng, B., Fan, Z.-P., & Ma, J. (2010). A method for partner selection of codevelopment alliances using individual and collaborative utilities. *International Journal of Production Economics*, *124*(1), 159–170. https://doi.org/10.1016/j.ijpe.2009.10.019

Ferrell, W., Ellis, K., Kaminsky, P., & Rainwater, C. (2020). Horizontal collaboration: Opportunities for improved logistics planning. *International Journal of Production Research*, *58*(14), 4267–4284. https://doi.org/10.1080/00207543.2019.1651457

Gansterer, M., & Hartl, R. F. (2018). Collaborative vehicle routing: A survey. *European Journal of Operational Research*, *268*(1), 1–12. https://doi.org/10.1016/j.ejor.2017.10.023

Gansterer, M., & Hartl, R. F. (2020). Shared resources in collaborative vehicle routing. *TOP*, *28*(1), 1–20. https://doi.org/10.1007/s11750-020-00541-6

Ghomi, V., Gligor, D., Shokoohyar, S., Alikhani, R., & Nezami, F. G. (2023). An optimization model for collaborative logistics among carriers in vehicle routing problems with cross-docking [Num Pages: 36 Place: Ponte Vedra Beach, United Kingdom Publisher: Emerald Group Publishing Limited]. *International Journal of Logistics Management*, *34*(6), 1700–1735. https://doi.org/10.1108/IJLM-11-2021-0515

Guajardo, M., & Rönnqvist, M. (2016). A review on cost allocation methods in collaborative transportation. *International Transactions in Operational Research*, *23*(3), 371–392. https://doi.org/10.1111/itor.12205

Hernández, S., & Peeta, S. (2014). A carrier collaboration problem for less-than-truckload carriers: Characteristics and carrier collaboration model. *Transportmetrica A: Transport Science*, *10*(4), 327–349. https://doi.org/10.1080/23249935.2013.766279

Karam, A., Reinau, K. H., & Østergaard, C. R. (2021). Horizontal collaboration in the freight transport sector: Barrier and decision-making frameworks. *European Transport Research Review*, *13*(1), 53. https://doi.org/10.1186/s12544-021-00512-3

Lai, M., Cai, X., & Hall, N. G. (2022). Cost Allocation for Less-Than-Truckload Collaboration via Shipper Consortium [Publisher: INFORMS]. *Transportation Science*, *56*(3), 585–611. https://doi.org/10.1287/trsc.2021.1066

Luan, J., Daina, N., Reinau, K. H., Sivakumar, A., & Polak, J. W. (2022). A data-based opportunity identification engine for collaborative freight logistics based on a trailer capacity graph. *Expert Systems with Applications*, *210*, 118494. https://doi.org/10.1016/j.eswa.2022.118494

McKinnon, A. C., & Ge, Y. (2006). The potential for reducing empty running by trucks: A retrospective analysis [Publisher: Emerald Group Publishing Limited]. *International Journal of Physical Distribution & Logistics Management*, *36*(5), 391–410. https://doi.org/10.1108/09600030610676268

Mohammed, R. A., Nadi, A., Tavasszy, L., & Bok, M. d. (2023). Data Fusion Approach to Identify Distribution Chain Segments in Freight Shipment Databases [Publisher: SAGE Publications Inc]. *Transportation Research Record*, *2677*(6), 310–323. https://doi.org/10.1177/03611981221147049

Naesens, K., Gelders, L., & Pintelon, L. (2009). A swift response framework for measuring the strategic fit for a horizontal collaborative initiative. *International Journal of Production Economics*, *121*(2), 550–561. https://doi.org/10.1016/j.ijpe.2007.04.004

Pan, S., Trentesaux, D., Ballot, E., & Huang, G. Q. (2019). Horizontal collaborative transport: Survey of solutions and practical implementation issues [Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00207543.2019.1574040]. *International Journal of Production Research*, *57*(15-16), 5340–5361. https://doi.org/10.1080/00207543.2019.1574040

Statista. (2024, January). Transportation emissions in the European Union - Statistics & Facts. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Road_freight_transport_by_journey_characteristics#Empty_runnings_of_road_freight_vehicles

Vargas, A., Patel, S., & Patel, D. (2018). Towards a Business Model Framework to Increase Collaboration in the Freight Industry [Number: 4 Publisher: Multidisciplinary Digital Publishing Institute]. *Logistics*, *2*(4), 22. https://doi.org/10.3390/logistics2040022

Wang, J., Yu, Y., & Tang, J. (2018). Compensation and profit distribution for cooperative green pickup and delivery problem. *Transportation Research Part B: Methodological*, *113*, 54–69. https://doi.org/10.1016/j.trb.2018.05.003

Wang, Y., Li, Q., Guan, X., Xu, M., Liu, Y., & Wang, H. (2021). Two-echelon collaborative multi-depot multi-period vehicle routing problem. *Expert Systems with Applications*, *167*, 114201. https://doi.org/10.1016/j.eswa.2020.114201

Xia, P., Zhang, L., & Li, F. (2015). Learning similarity with cosine similarity ensemble. *Information Sciences*, *307*, 39–52. https://doi.org/10.1016/j.ins.2015.02.024

TU Delft