Deliverable D 4.4

WP4 Report on identification of future innovation needs and recommendations for improvements

Project acronym:	RAILS
Starting date:	01/12/2019
Duration (in months):	43
Call (part) identifier:	H2020-S2R-OC-IPX-01-2019
Grant agreement no:	881782
Due date of deliverable:	Month 40
Actual submission date:	July 29^{th} 2023
Responsible/Author:	Rob Goverde (TU DELFT)
Dissemination level:	Public
Status:	Issued

Reviewed: no

Document history		
Revision	*Date*	*Description*
1	July 20th 2023	First issue for internal review
2	July 29th 2023	Second issue after internal review

Report contributors		
Name	Beneficiary Short Name	Details of contribution
Rob Goverde	TUDELFT	Coordinator
Ruifan Tang	LEEDS	Contributor
Zhiyuan Lin	LEEDS	Contributor
Ronghui Liu	LEEDS	Contributor

Funding

This project has received funding from the Shift2Rail Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement n. 881782 Rails. The JU receives support from the European Union's Horizon 2020 research and innovation program and the Shift2Rail JU members other than the Union.

Contents

Executive Summary 4

Abbreviations and acronyms 5

1. Objective 6

2. Introduction 7

3. Background 8

4. A Critical Examination of the Primary Delay Prediction Proof-of-Concept 12

4.1 Recent Advancements on Delay Prediction 12

4.1.1 Bird-eye View of the PoC Approach 13

4.2 A SWOT Analysis of the PoC 15

4.2.1 Strengths and Weaknesses 16

4.2.2 Opportunities and Threats 17

4.2.3 Recommendations from the PoC 18

5. A Critical Examination of the Incident Attribution Analysis Proof-of-Concept 20

5.1 Recent Advancements on Incident Attribution Analysis 20

5.2 A SWOT Analysis of the PoC 22

5.2.1 Strengths and Weaknesses 22

5.2.2 Opportunities and Threats 24

5.2.3 Recommendations from the PoC 25

6. Recommendations and Innovation Needs 27

6.1 Model Interpretability and Explainability 27

6.2 Gather Comprehensive and High-Quality Data 28

6.3 Choose Appropriate AI/ML or Hybrid Techniques 29

6.4 Ensure Scalability and Real-Time Capabilities 30

6.5 Collaborate with Domain Experts 31

7. Conclusions 33

Executive Summary

This Deliverable contains a critical examination of the work and of the results obtained in WP4 ‘AI for Traffic Planning and Management’, also against the current state-of-the-art in railways. From the proofs-of-concept and the experience gained by developing the tasks of the project, this document reports lessons learned, weaknesses and strengths shown by each exploited technology, technical and implementation recommendations, unaddressed issues, and innovation needs.

This deliverable provides a comprehensive analysis of the methodological and experimental Proof of Concepts (PoCs) conducted for the two case studies, ”Graph Embedding based Primary Delay Prediction” and ”Big Data on Incident Attribution Analysis”, as outlined in the previous WP4 deliverables.

For these two PoCs, the strengths are demonstrated by the exploited AI approach in efficiently processing complex network data and accurately predicting primary delays. However, potential weaknesses such as computational intensity during training and inference, data requirements, and model interpretability, will be properly identified as well. External opportunities to enhance the AI approach include integrating real-time data sources and collaborating with railway infrastructure providers and regulatory bodies for additional relevant data. Challenges and threats such as handling imbalanced datasets and ensuring general data security and privacy are essential to be considered.

To improve the effectiveness of the proposed AI approach, continuous monitoring and fine-tuning, scalable computing infrastructure, and explainable AI techniques are recommended. For broader applications, fostering collaborations, standardizing data formats, and exploring hybrid AI approaches are suggested to advance AI integration in the rail sector. This will lead to safer, more efficient, and reliable railway systems.

Abbreviations and acronyms

Abbreviations / Acronyms	Description
PoCs	Proof of Concepts
AI	Artificial Intelligence
WP	Work Package
SWOT	strengths, weaknesses, opportunities, threats
TPM	Traffic Planning and Management
SOTA	State-of-the-Art
SDNE	Structural Deep Network Embedding
PCA	Principle Component Analysis
ML	Machine Learning
DT	Decision Tree
RF	Random Forest
MLP	Multilayer Perceptron
SVD	singular value decomposition
LIME	Local Interpretable Model-agnostic Explanations
SHAP	SHapley Additive exPlanations
ORR	Office of Rail and Road
TRUST	Train Running Under System TOPS
TPE	TransPennine Express
GNN	Graph Neural Network
RL	Reinforcement Learning
ARIMA	Autoregressive Integrated Moving Average model
LSTM	Long short-Term Memory

1. Objective

This document aims to draw some conclusions from the work carried out in WP1 and WP4, helping to clearly identify possible future innovations, research directions, and impacts for the European railway sector. These objectives are strictly related to the work addressed in the previous WP4 deliverables, in which methodological and experimental Proofs-of-Concept (PoCs) have been carried out for two selected case studies, namely, ”Graph Embedding based Primary Delay Prediction” and ”Big Data on Incident Attribution Analysis”.

This deliverable’s primary objective, based on the findings from the Proof of Concepts (PoCs), is to suggest innovative strategies and recommendations that can facilitate the effective implementation of Artificial Intelligence (AI) in the domain of rail traffic planning and management. This document, accordingly, focuses on the following aims:

- Providing a succinct review of recent advancements related to the PoCs.
- Conducting a detailed SWOT analysis of the proposed PoCs with regard to the significant strengths (S) and underlying weaknesses (W) of the AI methodologies suggested, as well as identifying external opportunities (O) and potential threats (T) which may impact the technical execution.
- Identifying specific recommendations that may enhance the effectiveness of the proposed AI methodologies.
- Outlining broad recommendations for AI incorporation within railway planning and management. This primarily includes directives for further research, such as:
  - The further refinement of approaches, methodologies, models, technologies, and tools.
  - Expanding experimentation with additional data, case studies, pilot studies, and scenarios.
  - Applying these methodologies to other sectors and subsystems within the railway transport industry.

These recommendations will subsequently be used in WP5 to formulate strategic plans and roadmaps for AI integration within the rail sector.

2. Introduction

This deliverable provides a critical examination of the work and the results obtained in WP4 also against the current state-of-the-art in railways. It reports some technical/implementation recommendations and innovation needs that would support future investigations in the context of AI for railway maintenance and inspection.

The recommendations provided by this document can be subdivided into two main macro-categories:

1. Recommendations coming from the critical examination of the Proofs-of-Concept (PoCs) developed within the RAILS's WP4 which would be potentially useful to support future development of approaches, methods, models, technologies, and tools in the specific contexts of the PoCs and related areas.
2. General Recommendations, coming from lessons learned while both working at the PoCs and investigating the state-of-the-art of AI in railways, which aim at providing hints about practices and activities that would support the integration of AI across various railway applications.

The remainder of this deliverable is organised as follows. Chapter 3 summarises the findings of WP1 about the state-of-the-art and promising research directions in railway planning and management, as well as the documents and results produced during the project and addressing the topics investigated in the context of WP4, including the related scientific publications stemming from the project activities. Then, it provides the context and the background of the discussion reported in the present deliverable. Chapter 4 and Chapter 5 address critical examinations of the PoCs on "Graph Embedding based Primary Delay Prediction" and "Big Data on Incident Attribution Analysis", respectively. Chapter 4 and Chapter 5 share the same structure: Sections 4.1 and 5.1 discuss a high-level overview of the recent advancements in the context of the corresponding PoCs; Sections 4.2 and 5.2 present a bird-eye view of the investigative approaches and recall back some important technical details when they were implemented; While Sections 4.3 and 5.3 propose structured analyses of the implemented approaches in the form of SWOT (strengths, weaknesses, opportunities, and threats) analyses; lastly, Section 4.4 and 5.4 highlight the main recommendations resulted from the lessons learned while working at the PoCs. Then, Chapter 6 discusses the general recommendations and some innovation needs that would be required for the fast take-up of AI in railways. Lastly, Chapter 7 provides some concluding remarks.

3. Background

This section recalls most of the findings from the analyses carried out in the previous phases of the RAILS project with specific emphasis on AI applications for ”Traffic Planning and Management” (TPM) activities. Table 3.1 reports all the documents (deliverables and papers) resulting from the aforementioned research activities and specifies their main contributions/results.

Table 3.1: Published Documents discussing AI for Railway TPM Applications.

Focus	Document	Type	Main contribution(s)
Taxonomy	Deliverable D1.1: Definition of a Reference Taxonomy of AI in Railways [1]	PD	1. Delineation of a definition for AI in railway 2. Establishment of a taxonomy of AI in railway 3. Preliminary overview of regulations for AI 4. Identification of Railway Subdomains 5. Preliminary mapping of existing AI applications on Railway Subdomains
	Artificial Intelligence in Railway Transport: Taxonomy, Regulations, and Applications [2]	SP	
State of the Art	Deliverable D1.2: Summary of Existing Relevant Projects and State-of-the-Art of AI Application in Railways [3]	PD	1. Review of projects conducted worldwide (with emphasis on S2R projects) dealing with AI in Railway Subdomains 2. Review of scientific papers dealing with AI in Railway Subdomains 3. Preliminary definition of future direction towards the integration of AI
	A Literature Review of Artificial Intelligence Applications in Railway Systems [4]	SP	Extended review of scientific papers dealing with AI in Railway Subdomains
	A Systematic Review of Artificial Intelligence Public Datasets for Railway Applications [5]	SP	In-depth review of publicly available datasets for each Railway Subdomain
Application Areas	Deliverable 1.3: Application Areas [6]	PD	1. Identification of relevant railway Application Areas for AI together with the main challenges to be tackled for its effective integration basing on: i) The review of projects conducted worldwide and the scientific literature dealing with AI in railways; ii) Suggestions from the Advisory Board; and iii) the results from a comprehensive survey submitted to researchers and practitioners from different organisations operating worldwide 2. Delineation of basic AI usage guidelines to select the most appropriate AI approach by taking into account: i) the goal; ii) the type of available data; and iii) the required responsiveness of the AI system
Transferability	Deliverable 4.1: WP4 Report on Case Studies and Analysis of Transferability from Other Sectors [7]	PD	1. Review of AI-based emerging technologies developed in other transport (i.e., Aviation and Automotive) sectors other than railways 2. Identification of AI approaches that can be transferred to or potential to be adapted for railway applications
AI-based Graph Embedding for TPM	Predicting Primary Delay of Train Services Using Graph-Embedding Based Machine Learning [8]	SP	1. Incorporating both network spatial characteristics and historical delay into a train delay prediction framework 2. Implementing deep neural network-based graph embedding technique for extracting network features both globally and locally 3. For the first time, combining graph embedding approaches with matrix decomposition to generate route embedding vectors as an important feature for delay prediction.
PoCs Development	Deliverable 4.2: WP4 Report on AI approaches and models [9]	PD	Identification of PoCs to be developed together with Research Questions, Methodology, Reference Datasets, AI and ML Models, and Expected Results
	Deliverable 4.3: WP4 Report on experimentation, analysis, and discussion of results [10]	PD	Development of identified PoCs including model/architecture description, data generation, training and validation, and evaluation and discussion of results
Recommendations	Deliverable 4.4: WP4 Report on identification of future innovation needs and recommendations for improvements [this document]	PD	1. Identification of sectorialized recommendations oriented at supporting AI integration in WP4 PoCs' topic and related areas 2. Identification of general recommendations aiming at supporting AI integration across different railway applications

PD: Project Deliverable; SP: Scientific Paper

Herein, we recall some of the results obtained within the first phase of the project as they impacted the subsequent investigation carried out in WP4. First, a *Taxonomy of AI in Railway* has been introduced and seven *Railway Subdomains* have been identified for the RAILS investigation [1, 2]. Then, in Deliverable D1.2, for each of the subdomains, a review of the *State-of-the-Art (SOTA) of AI* in railways has been developed by analysing i) research projects conducted worldwide (with a particular focus on S2R projects) and ii) the scientific literature collected within our scope. The latter part has been then extended into a comprehensive Journal Paper [4]. The main statistical distributions regarding the investigated scientific paper and research projects we generated from the SOTA are reported in Fig. 3.1. As we can conclude, as the second most discussed area of the railway sector, TPM topics account for a quarter of the investigated papers and about 16% reviewed projects. Fig 3.2 summarises the information about what percentages of the investigated papers fall into each TPM topic.

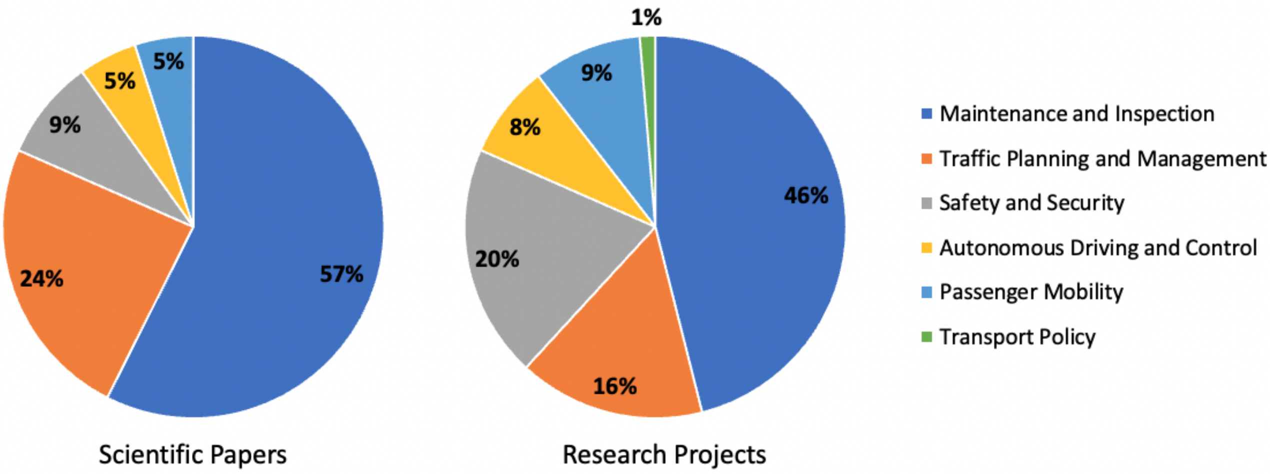

Fig. 3.1. Distribution of Scientific Papers and Research Projects per Railway Subdomain.

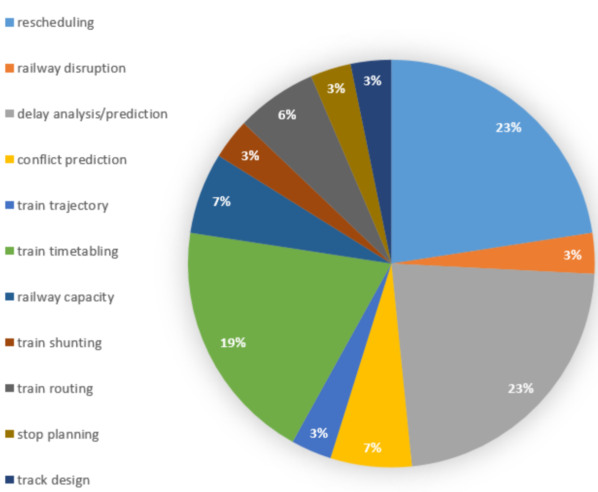

Fig. 3.2. Distribution of Scientific Papers per TPM topic.

- **Rescheduling** are the studies that consider changing the timing or routing of a train's journey due to factors such as track maintenance, weather conditions, operational disruptions, or to accommodate passenger needs.

- **Railway disruption** involves the strategies oriented at minimising the impact of unexpected events, such as equipment failure, track obstructions, or extreme weather conditions, on train operations to maintain regular schedules and passenger satisfaction.
- **Delay analysis/prediction** refers to the utilisation of data analytics and predictive modelling to estimate the likelihood and extent of train delays, thereby enabling proactive responses to potential disruptions.
- **Conflict prediction** investigates the process of identifying and forecasting potential conflicts or collisions between trains on shared tracks or intersections, using algorithms and data analysis to ensure efficient scheduling and safety.
- **Train trajectory** papers relate to the analysis regarding the path that a train takes from its starting point to its destination, considering factors such as speed/distance, timing, safety, energy consumption rate, and performance of railway operations.
- **Train timetabling** is the most commonly discussed topic that plans train arrivals, departures, and stops at various stations, considering factors such as passenger demand, train speed, and track capacity, to ensure efficient, punctual, and reliable railway operations.
- **Railway capacity** papers explore the maximum number of trains that can safely and efficiently run on a given rail network or line within a specific period, considering factors such as track layout, signalling systems, speed restrictions, and station stops.
- **Train shunting** focuses on rearranging railway vehicles within a train yard or between different tracks- where train cars are switched, sorted, and assembled into new trains - to improve efficiency, safety, and reduce operational costs.
- **Train routing** papers determine the best route for a train to travel from its origin to its destination, taking into account factors like track availability and operational efficiency.
- **Stop planning** refers to the process of determining where and when a train should stop along its route, taking into account factors like passenger demand, station capacity, and overall schedule efficiency.
- **Track design** involves planning and engineering the layout of railway tracks, considering factors such as train speed, safety requirements, geographic terrain, and infrastructure needs, to ensure efficient and safe railway operations.

Table 3.2 summarises the main AI approaches researchers and practitioners have leveraged subdivided for TPM tasks. Notably, here we incorporated all the TPM topics into four categories: Strategical Planning (including stop planning and track design); Tactical Planning (including train timetabling, railway capacity, train shunting, and train routing); Traffic Analysis (including delay analysis/prediction, conflict prediction, and train trajectory); Rescheduling and Disruptions (including rescheduling and railway disruption).

Table 3.2: Mapping AI Algorithms/Models with TPM Tasks.

TPM Task	AI Algorithm / Model																						
	Tree-based				*Regression*			*Clustering*		*others*										*EC*			
	DT	RF	GBDT	EGBT	LR	SWLR	GPRRQ	EM	K-means	AdaBoost	XGBoost	LDA	QDA	FRA	FAHP	KRB	ADP	RL	HHMM	GA	PSO	SI	ACO
Strategical Planning			x																				
Tactical Planning																		x		x		x	
Traffic Analysis	x	x	x		x			x		x				x							x		
Rescheduling and Disruptions																		x			x		

Primary Header Acronyms: Evolutionary Computation (EC)
Secondary Header Acronyms: Decision Trees (DT), Random Forest (RF), Gradient Bosting Decision Tree algorithm (GBDT), Ensemble Bagged Trees (EBGT), Linear Regression (LR), Stepwise Linear Regression (SWLR), Rational Quadratic Gaussian Process Regression (GPRRQ), Expectation-Maximization Algorithm (EM), K-means algorithm (K-means), Adaptive Boosting (AdaBoost), eXtreme Gradient Boosting (XGBoost), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Fuzzy Reasoning Approach (FRA), Fuzzy Analytical Hierarchy decision-making Process (FAHP), Knowledge Rules-Based (KRB), Approximate Dynamic Programming (ADP), Reinforcement Learning (RL), Hierarchical Hidden Markov Models (HHMM), Genetic Algorithms (GA), Particle Swarm Optimization (PSO), Swarm Intelligence (SI), Ant Colony Optimization(ACO).

The state-of-the-art analysis conducted earlier set the groundwork for more in-depth studies aimed at recognizing key railway application areas outlined in Deliverable D1.3. These are groups of railway applications that could potentially benefit from artificial intelligence (AI). Our findings from these explorations, along with the analysis of AI methodologies proposed in other transport sectors (referenced in Deliverable D4.1), aided in pinpointing and cultivating the Proof of Concepts (PoCs) in WP4, detailed in Deliverables 4.2 and 4.3. To summarize, the research carried out within the mentioned deliverables and all other documents referenced in Table 3.1 led to the creation of practical recommendations, which will be further elaborated in subsequent sections of this report.

4. A Critical Examination of the Primary Delay Prediction Proof-of-Concept

In this chapter, we present an analysis of a Proof-of-Concept (PoC) conducted under the topic of ‘Graph Embedding based Railway Primary Delay Prediction’, with a specific focus on the rail transportation domain. The primary objective of the PoC was to explore and evaluate the effectiveness of delay prediction frameworks using advanced AI algorithms, such as the Structural Deep Network Embedding (SDNE) algorithm and Principle Component Analysis (PCA). The successful implementation of the PoC highlights the potential for data-driven approaches in enhancing delay management and operational efficiency in various industries. Throughout the PoC, the SDNE algorithm played a crucial role in maintaining the structural relationships between nodes in the complex network. By effectively representing related nodes in close proximity within a lower-dimensional vector space, SDNE demonstrated its capability to capture and preserve essential network connections and dependencies. This, in turn, led to improved accuracy in delay predictions, enabling more proactive and informed decision-making for managing passenger service punctuality.

In order to assess the performance of the implemented PoC, we will conduct a comprehensive SWOT analysis. By examining the strengths (S), weaknesses (W), opportunities (O), and threats (T) of the delay prediction framework utilized in the PoC, we aim to gain deeper insights into its potential for broader application and areas for further improvement. Moving forward, the findings from the PoC and the SWOT analysis will serve as valuable inputs for future research endeavours in the domain of delay prediction frameworks. By identifying areas of strengths and opportunities for enhancement, we can propose general recommendations not only to refine this particular delay prediction model but also to give some promising research directions for the future implementation of delay prediction frameworks. Moreover, these recommendations will facilitate the development of more robust and scalable solutions for predicting delays in a more complex railway network, fostering data-driven decision-making across diverse industries. The following sections of this chapter will be elaborated with the details of the conducted PoC, present the SWOT analysis results, and outline general recommendations for advancing delay prediction frameworks. Through this comprehensive assessment, the objective is to contribute valuable insights to the field of AI-based primary delay prediction for railway services by exploring the effectiveness of advanced algorithms like the Structural Deep Network Embedding (SDNE) and conducting a detailed SWOT analysis towards the proposed method, with the final aim of enhancing the accuracy and efficiency of delay prediction frameworks.

4.1. Recent Advancements on Delay Prediction

As we described before, in the preceding deliverables D4.1[7], D4.2[9] and D4.3[10], we identified potential AI approaches that can be transferred to or be adapted for TPM applications, and then several research questions, the aim of study, methodology, reference datasets, ML models/architecture description, data generation, training and validation, and evaluation and discussion of results have been introduced, respectively. The aim of this case study is to predict the overall degree of primary delay level for individual train services in the future. This estimation is based on historical data from different time periods in the railway's operation. The analysis takes into account the static characteristics of each station where trains pass by or dwell, as well as the structural network characteristics, including the connectivity between stations, link travel times, and network density in various areas.

To achieve this, we utilise the Structural Deep Network Embedding (SDNE) algorithm, this graph embedding algorithm was first created as an effective dimensionality reduction tool in the computer science field by Wang et al. [11]. The main idea behind SDNE is to keep related nodes closer to one other in vector space so that the original network's structural relationships can be preserved. In this case study, we have refined and enhanced the SDNE algorithm to interpret station dependencies and structural correlations. The process involves constructing a similarity network for a set of D-dimensional nodes, considering their neighbourhood information. Subsequently, each node in the graph is embedded into a d-dimensional vector space, where $d \ll D$. The main objective of this embedding is to ensure that related nodes are closer to each other in the vector space, thereby preserving the original network's structural relationships. Specifically, this concept behind 'embedding' is to create a vector for each railway station, with each element representing the scalar value on a specific vector direction in Euclidean space. Each value in the vector has no discernible significance, yet it does represent a characteristic of a certain station in part. When we wish to compare how similar two stations are, such a representation comes in handy. By doing this, we capture essential information about the station in a condensed form, facilitating analysis and predictions. Furthermore, we propose a technique to combine the obtained hypernode embedding vectors, which represent nodes or stations along a particular route. We create a route embedding vector that encapsulates and aggregates more structural information about the target railway network. This process effectively reduces the dimensionality of available features, making the analysis more efficient and insightful. In this regard, the SDNE approach considerably compresses the fundamental information, making vector operations simpler and faster than traditional mathematical procedures. The following requirements must be met by the expected route embedding depictions:

- Regardless of the length of a specific route, the obtained route embedding vectors must be uniform in size – this makes them more convenient to use as input features for subsequent prediction tasks.
- The route representations can explicitly reflect the characteristics of the entire route, including the density of en-route station cluster, the sequence of these stations, and the degree of congestion on this route.
- Local and global characteristics can be effectively preserved by route embedding vectors.

4.1.1. Bird-eye View of the PoC Approach

The primary objective of the entire methodology framework is to create a structural deep network representation for effective modelling of the highly non-linear structure present in the railway network. To achieve this, a unique deep learning model called the "Structural Deep Network Embedding approach" was proposed, drawing inspiration from successful applications of deep learning methods presented in prior works [11] and demonstrated to possess strong representational capabilities across various data types [12–15]. Notably, this study pioneers the application of such an approach to a public transit system, specifically within a train system.

Figure 4.1 illustrates the suggested SDNE framework, which takes the original railway network characteristics as input for the encoder-decoder layers. These layers incorporate the definition of first-order and second-order proximity and identify the connectivity status between any two nodes within the network. During the training process, the resulting embedding vector is updated to minimise the overall loss costs using specific loss functions corresponding to each proximity in the output layer. This optimisation process ensures that the encoder-decoder layers obtain the most suitable parameters. Consequently, each node within the railway network is assigned its final low-dimensional embedding representation, enabling a comprehensive and efficient structural deep network representation.

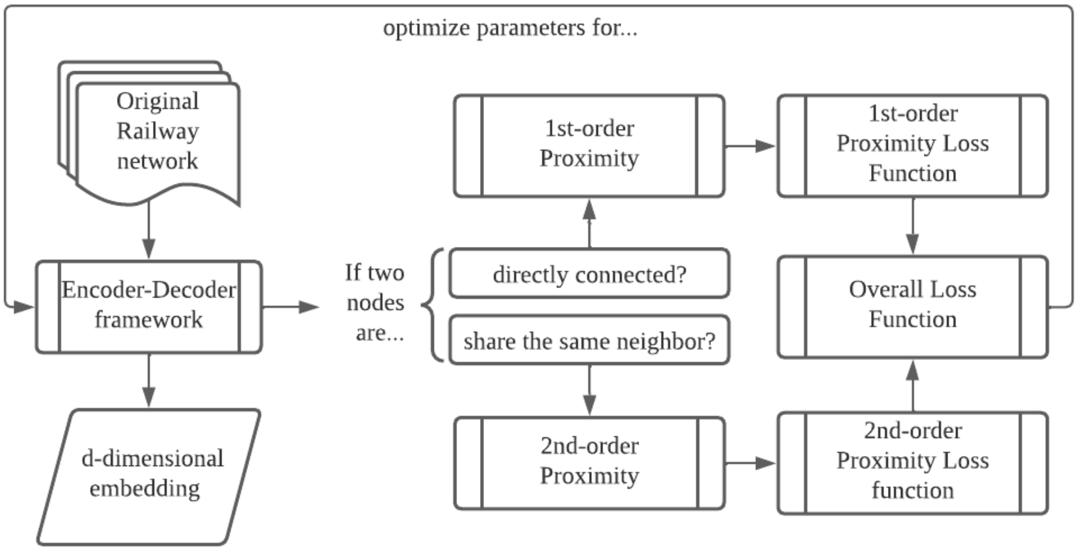

Fig. 4.1. The flow-charted SDNE framework

The flowchart (shown in Figure 4.2) represents an experimental system for predicting delays in the railway network. The designed experiment takes in data from a Network Topology Structure and splits it into two branches. The first branch uses the well-known Principle Component Analysis (PCA) algorithm [16] to compress the route vectors for services and then merges this data with timetable, operating train profile, and infrastructure data. This merged data is fed into three different machine learning predictors: Decision Tree (DT) [17], Random Forest (RF) [18], and Multilayer Perceptron (MLP) [19] for train delay level prediction.

The second branch uses SDNE to generate node embeddings for stations, which are then processed by SVD [20] to generate route embeddings for services. This branch also merges the timetable, operating train profile, and infrastructure data, and feeds the merged data into the same three machine learning predictors as the first branch. The reason why we choose these three benchmarks is that they are well-established algorithms that have been extensively tested and validated, and we want to obtain a standard of performance under the scope of the defined objevctives and research questions in D4.2 [9]. The machine learning predictors in both branches are used to predict delays for each train service. Overall, this system utilizes both PCA and SDNE to process data and combines the resulting data with timetable, operating train profile, and infrastructure data to make predictions using machine learning.

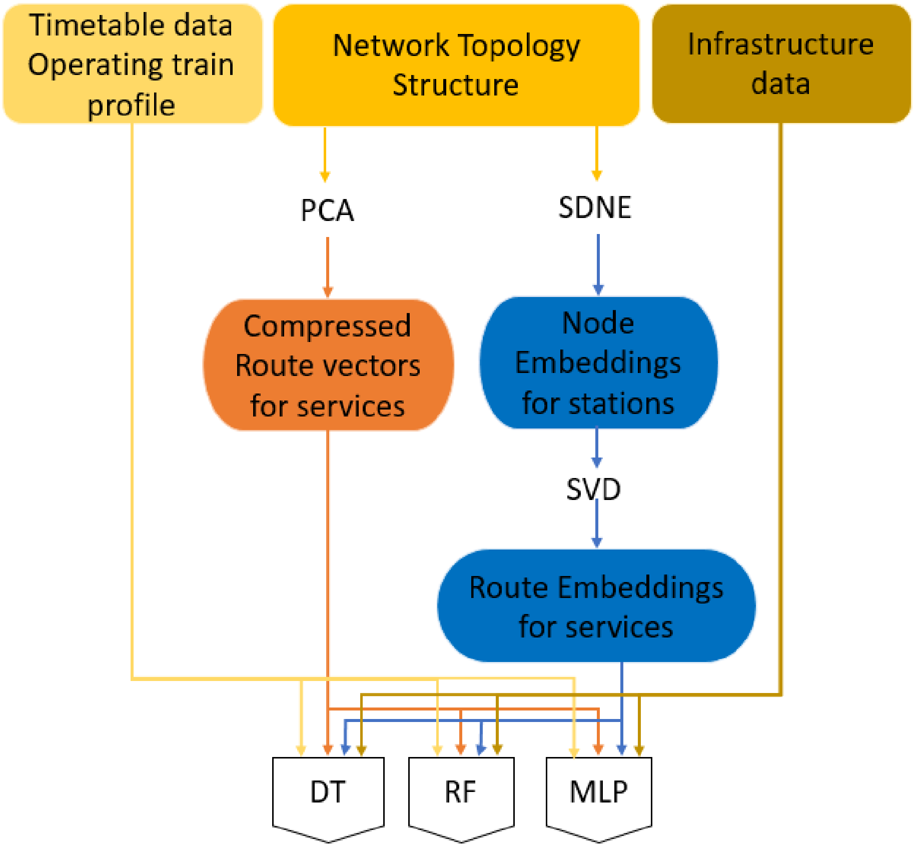

Fig. 4.2. Experimental design for Primary Delay Prediction

4.2. A SWOT Analysis of the PoC

In this segment, we bring to the forefront the inherent Strengths (S) and Weaknesses (W) of the methodology we have examined, while pinpointing external Opportunities (O) and Threats (T) that could respectively bolster or hinder its technical realization. It is crucial to note that this SWOT analysis isn't geared towards determining the market feasibility of the solution we've explored. It is rather used as an assessment approach regarding the proposed method. All pertinent aspects identified through our research and trials are outlined in Fig. 4.3, arranged following the SWOT format. Here is a SWOT analysis for integrating the Structural Deep Network Embedding (SDNE) model into railway delay prediction.

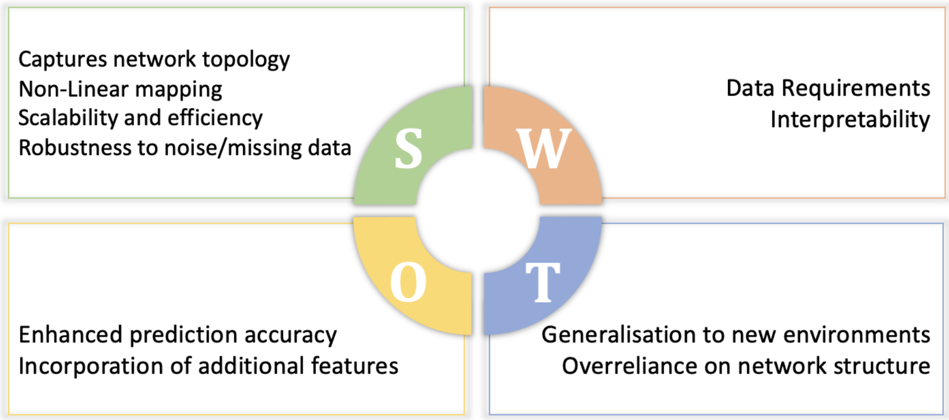

Fig. 4.3. Relevant aspects from the Primary Delay Prediction PoC arranged according to the SWOT Structure

4.2.1. Strengths and Weaknesses

Captures network topology: This states the SDNE's capability of effectively capturing and interpreting the structural and topological features of railway networks. It considers both spatial (geographical) and connectivity (inter-relations between different nodes or stations) patterns in these networks. By understanding the layout of the network and how different parts of the network connect and interact with each other, the model can identify patterns and trends that might not be immediately apparent. This makes it possible to leverage these inherent network characteristics to enhance the accuracy of its delay predictions, by providing a comprehensive overview of the vehicle interactions/geographical correlations between services over the entire railway system.

Non-linear mapping: Railway networks often involve complex relationships and dependencies that are not directly proportional, or non-linear. This might include the relationship between the number of trains on a track and the corresponding delays, or the interplay between different components of the network. SDNE has the ability to model these non-linear relationships, meaning it can effectively capture and represent these complex interactions and dependencies in the network. This is particularly advantageous for predicting delays, which often arise due to these intricate interactions between different components of the railway network.

Scalability and efficiency: As railway networks often involve large-scale systems with numerous stations, tracks, and connections, any effective model needs to be able to handle and process a vast amount of data. SDNE's semi-supervised training process has been designed to be scalable and efficient, meaning it can manage large-scale railway networks without compromising on speed or accuracy. This allows the model to learn from a wealth of data, continually improving its predictions and insights over time. This scalability makes SDNE suitable for real-world railway systems, which are often complex and extensive.

Robustness to noise and missing data: One of the challenges of working with real-world data is that it is often imperfect. Data may include noise (errors or random fluctuations) or may be incomplete due to various reasons. SDNE, however, has been designed to handle such challenges. It is robust to noise and can work with incomplete data, still managing to generate reliable predictions. This feature is particularly useful in the context of delay prediction, where the data may contain inconsistencies or missing information, allowing the model to provide reliable insights even when faced with imperfect data.

High Volumn data requirement: Though the SDNE model is proved to be robust to noise or outlier data, the functionality and effectiveness of the SDNE model largely depend on the availability of substantial, comprehensive data. It requires comprehensive and accurate data about the network, including historical delays, train schedules, and network topology, for training and learning purposes. Gathering such a wide range and volume of data can pose significant challenges, especially in the context of older railway systems that have not been fully digitized. These systems may not have the necessary technology or resources to collect and store the extensive data required. In cases where the data is insufficient, incomplete, inaccurate, or difficult to be collected, the efficiency and accuracy of the SDNE model can be negatively impacted. It may not be able to fully learn the necessary patterns and trends, which can limit its effectiveness in predicting delays and analyzing the railway network.

Model interpretability: While the SDNE model is effective in generating representations of the railway network and predicting delays, these representations may not always be easily interpretable. The complex, high-dimensional nature of the representations can make it difficult to understand which specific factors or features contribute to the delay predictions. Even though the model can identify patterns and make predictions based on the data, understanding why it makes certain predictions can be challenging, especially for those who do not have much expertise in ML implementation. The lack of transparency and interpretability can make it hard to provide clear, detailed explanations or actionable insights based on the model alone. It could also make it more difficult to diagnose any issues with the model or to convince stakeholders of the value and reliability of the model's predictions.

4.2.2. Opportunities and Threats

Enhanced prediction accuracy: Utilising the SDNE within railway delay prediction models creates a significant opportunity to augment prediction accuracy. SDNE's inherent capabilities, such as understanding complex relationships and structural information within railway networks, can lead to the generation of more precise delay predictions. This higher level of accuracy can streamline operations, minimise disruption, and enhance service levels for passengers. Better predictions can also facilitate more effective resource allocation and improve overall system management, providing significant benefits across the entire railway operation.

Incorporation of additional features: SDNE's flexibility in handling additional features or data points during the learning process is another key advantage. It opens up avenues to include more nuanced data, such as weather conditions that can impact railway operations, historical performance data for benchmarking and trend analysis, or maintenance schedules that can affect service availability. The inclusion of these diverse datasets could significantly boost the model's prediction capabilities, leading to a more well-rounded, insightful, and effective model for railway delay prediction.

Generalization to new environments: Although SDNE's versatility and adaptability have been demonstrated across various fields, challenges may arise when applying it to railway systems with distinct characteristics or those operating under unique regional conditions. Factors such as differing operational procedures, infrastructure variations, or localised regulations might pose barriers to the general applicability of the SDNE model. Addressing these issues may require meticulous adaptation, potentially involving retraining the model with region-specific data or modifying it to better account for local conditions, thereby ensuring its effectiveness across diverse environments.

Overreliance on network structure: While SDNE's strength lies in its ability to capture network topology, its heavy reliance on this aspect could potentially limit its comprehensiveness. Factors contributing to railway delays extend beyond network structure and include variables like operational disruptions, scheduled and unscheduled maintenance activities, and external events like weather conditions or accidents. These factors may not be explicitly considered in the SDNE model, thus possibly limiting its predictive accuracy in certain contexts. To ensure a comprehensive prediction model, it might be necessary to integrate SDNE with other tools or models that account for these external variables, thus creating a more robust and accurate predictive system.

4.2.3. Recommendations from the PoC

Interpretable AI Techniques: The challenge with the interpretability of the SDNE model can be addressed by focusing on incorporating techniques that enhance interpretability. This would involve developing methods or layers that can help with illustrating the model learning and decision-making process. Efforts should be made to understand how the model's encoding-decoding layers are processing information and making connections between data points, thus resulting in specific delay predictions. To achieve this, one may consider integrating model-agnostic methods such as LIME (Local Interpretable Model-agnostic Explanations) or SHAP (SHapley Additive exPlanations) which help in providing insights into what the model has learned. Alternatively, certain model-specific techniques, such as 'attention mechanisms', can also be employed to shed light on the important features or relationships that the model is focusing on while making predictions. Making the SDNE model more interpretable will not only make the results more understandable but also enhance trust in the model, making its predictions more actionable for the end users.

Feature Incorporation: While SDNE has the inherent ability to incorporate additional features into the learning process, it is crucial to leverage this flexibility effectively. Relevant features that hold potential to enhance the prediction capabilities of the model should be identified and incorporated. This could include a diverse range of features such as weather conditions, historical performance metrics, and maintenance schedules. However, simply incorporating these features is not enough. It is equally important to understand their specific influence on delays. For instance, determining how various weather conditions affect train delays can inform how the model should weigh these features. Understanding historical performance can also shed light on recurrent issues or patterns that contribute to delays, while maintenance schedules can offer insights into the times when delays are most likely due to planned work. Each of these features should be studied in-depth and their correlation with delays should be understood, in order to effectively train the model to consider these aspects in its predictions.

Customization and Adaptation: Given that the SDNE model may be applied across diverse contexts or railway systems, the ability to customise and adapt the model becomes paramount. It should be designed and developed with an understanding that a one-size-fits-all approach may not be effective in different geographical locations or operational contexts. For instance, a model trained on a railway system in a densely populated urban area may not perform as well when applied to a railway system in a rural area due to different patterns of usage, network configurations, and operational challenges. To address this, the model's structure should be flexible enough to accommodate variations in the input data and the patterns it needs to learn. This could involve building tunable parameters into the model that can be adjusted based on the specific context it is being applied to, or designing the model in such a way that it can be easily retrained on new data when necessary. Also, the model's transferability - its ability to apply learnings from one context to another - should be continually evaluated and improved. This ensures that the model remains effective and reliable, regardless of the specific railway system it is being applied to.

5. A Critical Examination of the Incident Attribution Analysis Proof-of-Concept

The British Rail Delivery Group's Delay Attribution Report to the Office of Rail and Road (ORR)[1] presents several areas for improvement to enhance the quality, reliability, and understanding of delay attribution data. The July 2019 scoping stage report[2] identified ten recommendations and assigned responsible owners to address them. Despite progress in certain areas, several recommendations remain unexplored. To stimulate advancement in delay attribution analysis and respond to the Steering Group's vision, we aim to further automate the attribution process of cascading delays. The goal is to explore how AI techniques can be effectively integrated into this process.

A deeper understanding of how delays at specific locations impact the broader network is crucial for both infrastructure providers and train operators. The existing TRUST system, which provides delay attribution data, only reviews trains delayed by at least 3 minutes. Delays less than this threshold are automatically attributed to the responsible railway company and Network Rail without an in-depth investigation into the root causes of the delays.

Several complex factors, such as timetable conflicts and track access rights, determine the duration and range of delay propagation across the network. These nonlinear spatial-temporal interactions can be challenging to predict accurately using traditional methods. Similarly, different railway disruptions and abnormal events can be triggered by various determinants, some of which share the same root causes while others do not. Employing conventional statistical analysis or descriptive models to analyse all observed relations or determinants may not yield the correct delay propagation chain.

The study we propose to undertake has two primary objectives. First, we aim to use Big Data techniques to visualise historic train delay records interactively, allowing us to reproduce how delays were triggered and subsequently propagated due to small disturbances, disruptions, or unexpected events. Second, we want to understand how these disturbances evolve into observed primary delays and then propagate along specific lines/routes of the network. By learning from these patterns, we aim to predict whether a delay will occur or propagate between particular locations, time points, and train services.

In this endeavour, we will use Big Data for interactive delay attribution visualisation and Graph Neural Network techniques for predicting potential propagation links. We intend to train a link prediction model that can predict whether a propagation link should exist between two nodes, enhancing our understanding of delay causation and propagation.

5.1. Recent Advancements on Incident Attribution Analysis

In our case study, we got more interest in analysing the delay attribution data, as it is possible to predict the propagation of delays and the occurrence of secondary delays. For example, if a delay event is caused by infrastructure issues, it may result in train bunching, which

[1] https://www.orr.gov.uk/sites/default/files/2020-09/rdg-delay-attribution-review-report-2020-09-28.pdf

[2] https://www.orr.gov.uk/sites/default/files/2021-06/delay-attribution-review-scoping-stage-report.pdf

can cause delays to other trains on the same route. By analysing historical delay data, it is possible to identify these patterns and predict the likelihood of secondary delays occurring in the future. This information can then be used to develop strategies to minimise the impact of delays on train services.

Understanding the root causes of performance issues is not that easy, due to the fact that the railway system has complex interactions and dependencies between individual components (i.e., passengers, trains, staff, stations, timetables, junctions, weather). Secondly, the propagation of delays is sensitive to small variations in inputs that can cause an escalating chain of events, such as cascading delays across the network. In addition, an observed delay can be affected or determined by rare combinations of events.

Our proposed tools consist of a set of interactive visualisations to explore the complex interactions between modelled train services and events. Based on this, a GraphSAGE-based model has been developed to estimate the potential primary/secondary delay resulting from the existing incidents/train service event across the network of TPE routes[3]. In addition, a pilot intervention simulation has been performed with several supervised machine learning techniques, with the purpose of improving overall service quality, see Fig 5.1.

Fig. 5.1. High-Level Architecture for Big Data on Incident Attribution Analysis

3-D interactive visualizations This module aims to simulate how the sequential chain reaction is triggered between different incidents and trains, as well as between trains themselves, in an informative space of a hybrid spatial-temporal scale. We will inspect the evolutionary process of how a ”significant” delay develops from small disturbances to an observable primary delay and then secondary delays, in a more intuitive and clear way. Consequently, how these delays subsequently affect the punctuality of other train services. Multiple essential information will be effectively illustrated such as the length of delay minutes, scale of incident/delay, the cause of the incident, triggering relationships between delays, and the significance of dependencies between services.

With the use of such informative visualisation, over thousands of statistical values, such as places, trains or times, can be easily displayed and understood. Interactions allow the user to find out more information or compute on-demand statistics for particularly interesting scenarios. Our visual summaries not only provide insights of

[3] https://www.tpexpress.co.uk/

problematic train services and locations, but also enable users to delve further into the information to comprehend the causes of these delays and aid in the planning of intervention policies.

Intervention simulation Once the potential reasons for the reactive delay have been determined, interventions that aim to shorten these delays might be suggested. The modelling and visualisation tools can then be used to recreate the sequential occurrence process of events with a set of input data that describes what the interventions are intended to accomplish in order to validate the efficacy of these interventions, such as reducing or preventing the causes of significant delays that the railway stakeholders might be interested in resolving. For example, reducing the number of track-based primary incidents, or reducing a range of incident durations.

GraphSAGE-based model In this module, our task is to learn if an edge exists between a provided node (service) and the existing nodes (services) we represented in the first module. In other words, exploring the possible responsible train and the potential reacted train for a newly introduced train service in the network that is characterised by an analysis of Network Rail attributed delay data. We use our implementation of the GraphSAGE algorithm [21] to build a model that predicts propagation links in our proposed TPE-Network Rail hybrid dataset. This problem is treated as a supervised link prediction problem on a heterogeneous delay propagation network with nodes representing incident and delay cases for train services.

5.2. A SWOT Analysis of the PoC

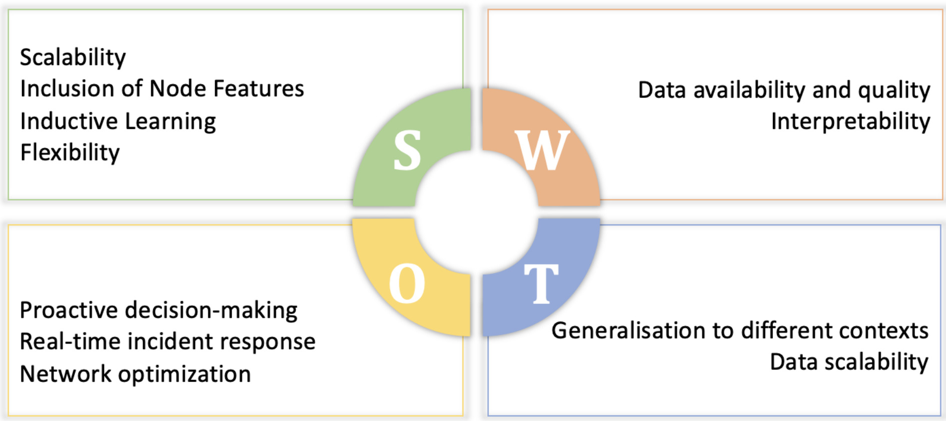

Fig. 5.2. Relevant aspects from the Incident Attribution Analysis PoC arranged according to the SWOT Structure

5.2.1. Strengths and Weaknesses

Scalability: One of the key advantages of GraphSAGE is its capability to scale efficiently with the size of graphs, making it highly effective for dealing with large-scale systems like railway networks. Its unique approach, which involves the sampling and aggregation of information from the local neighbourhoods of each node, allows it to handle graphs with millions of nodes or more without losing performance. In the context of large railway networks, this scalability is crucial. With thousands of stations, routes, and intersections to consider, GraphSAGE can quickly process this extensive information, thereby saving on computational resources, reducing processing time, and ensuring timely and accurate insights that can be utilised for effective decision-making and efficient network operation.

Incorporation of Node Features: GraphSAGE sets itself apart from other models like DeepWalk and Node2Vec by its ability to incorporate node feature information in addition to the structural properties of the graph. Each node in a railway network might have unique attributes such as station size, frequency of trains, connection types, and more. By incorporating these attributes, GraphSAGE is able to generate more detailed, informative, and discriminating embeddings that capture both the local and global structure of the network. This nuanced understanding can lead to richer insights, improve the accuracy of delay predictions, and offer a more comprehensive view of the network's functioning.

Inductive Learning: GraphSAGE is equipped with the capability for inductive learning, meaning it can learn from the existing data and apply this learning to unseen or new nodes and graphs. As railway networks are dynamic and continually evolving entities, new stations can be added, old ones can be decommissioned, and routes can be altered. In such cases, GraphSAGE's ability to generalise from learned patterns to unseen data is invaluable. It can predict the impact of these changes and adapt its model to account for them, ensuring consistent and accurate performance even in a changing network environment.

Flexibility: Another distinct advantage of GraphSAGE is its inherent flexibility, allowing the model to be customised based on specific requirements or tasks. It supports a variety of aggregation functions, accommodating different ways of summarising neighbourhood information. Moreover, it can be tailored to different types of graphs and prediction tasks, making it a versatile tool for diverse railway systems. Whether the task at hand is delay prediction, route optimisation, or network analysis, GraphSAGE can be adjusted to cater to these specific needs. This flexibility enhances the model's applicability and ensures it provides relevant and useful insights regardless of the particular railway system or task.

Data availability and quality: The effectiveness of GraphSAGE is intimately linked to the availability and quality of data on railway incidents and their cause-effect relations. Adequate, high-quality data is critical for the model to learn accurate representations and make dependable predictions. In railway networks, there are many potential sources of data, including train schedules, delay records, maintenance logs, and incident reports. However, these sources can sometimes be incomplete or inconsistent, with gaps in the data or discrepancies between different records. For instance, information about a particular incident might be missing, or cause-effect relationships might be incompletely recorded. These issues can limit GraphSAGE's ability to learn accurate representations of the network and may result in less reliable predictions. Therefore, steps must be taken to ensure that the data used for training and prediction is as complete and accurate as possible.

Interpretability: Another challenge with GraphSAGE, as with many machine learning models, is the difficulty of interpreting its output. While GraphSAGE can generate highly informative embeddings that capture a lot of detail about the railway network, understanding these embeddings and the specific factors that contribute to incident prediction and cause-effect relationships can be challenging. The model functions in a 'black box' manner, meaning it's not immediately clear how it makes its decisions or predictions. This can make it difficult to provide clear, understandable explanations for its predictions, which is especially problematic when those predictions need to be justified or explained to stakeholders. This lack of interpretability can also limit the ability to gain deeper insights into the underlying patterns in the data, and to use those insights to inform further improvements or interventions. Therefore, supplementary techniques, such as feature importance analysis or model explanation tools, might be necessary to improve the interpretability of GraphSAGE's predictions.

5.2.2. Opportunities and Threats

Proactive decision-making: GraphSAGE's ability to accurately predict cause-effect relationships can play a significant role in promoting proactive decision-making in railway operations. Railway authorities can, by interpreting these cause-effect relationships, identify potential consequences of various incidents and predict their possible impacts. For instance, understanding that a particular mechanical failure can lead to significant delays allows operators to preemptively allocate resources to mitigate the issue or reroute traffic. This anticipatory action can help minimize disruptions, optimize resource allocation, and ultimately, improve overall operational efficiency. However, the effectiveness of such proactive measures will depend on the quality of the predictions and the ability of the authorities to act on these insights in real-time.

Real-time incident response: GraphSAGE offers the potential to facilitate real-time incident response systems in railway operations. By continuously monitoring, analyzing, and predicting cause-effect relationships, the model can provide valuable insights that enable operators to respond promptly to incidents. For example, in the event of a sudden track failure, the system could quickly predict its impact on the railway network and suggest appropriate response strategies, such as rerouting trains or adjusting schedules. Rapid implementation of these strategies could lead to reduced delays, enhanced passenger safety, and improved customer satisfaction. However, developing such real-time systems would require integrating GraphSAGE with other operational systems and processes, which can be technically challenging.

Network optimization: By modelling and understanding the cause-effect relationships and dependencies within the railway network, GraphSAGE can assist in identifying critical points in the infrastructure, such as heavily used nodes, weak points prone to failures, and bottlenecks that constrain capacity. These insights can guide infrastructure planning and capacity management, helping to optimise the network's configuration and operation for improved performance and reliability. For instance, knowledge about recurrent delays at a particular station due to high traffic could inform decisions to increase capacity or optimise scheduling at that location. However, translating these insights into effective interventions would require careful planning and execution, as well as consideration of various practical constraints and trade-offs.

Generalisation to diverse railway systems: Railway systems can vary significantly in terms of infrastructure, operational practices, and incident characteristics, among other factors. While GraphSAGE can learn from one system and apply its insights to others, its ability to generalise across diverse railway environments may be limited. Some local nuances or unique features may not be adequately captured by the model, affecting its predictive performance. Adapting the model to different railway systems and ensuring that it continues to perform robustly across different contexts is a challenging task. This might require additional data collection, model fine-tuning, or even retraining the model using local data. The task of converting raw railway incident data into a suitable graph representation is complex and requires careful consideration. Decisions need to be made about how to represent nodes and edges, what features to include, and how to structure the graph to best capture the underlying cause-effect relationships. For instance, a node could represent a station or a particular type of incident, while an edge could represent a causal relationship or a sequence of incidents. The choice of features could also impact the model's performance - for instance, including information about the time of an incident or the type of train involved might improve the model's predictive accuracy. However, making these decisions requires a deep understanding of both the railway system and the GraphSAGE model, as well as a thoughtful consideration of the implications of different choices.

Scalability to large-scale railway networks: Railway networks can be large and complex, with a vast amount of incidents and cause-effect relations. While GraphSAGE is designed to be scalable, handling large-scale networks can still present challenges in terms of computational efficiency and memory requirements. The model needs to process and learn from extensive data, which can require significant computational resources and lead to long training and inference times. Strategies to ensure scalability, such as distributed processing, efficient data sampling, or graph partitioning, might be necessary to enable the use of GraphSAGE in large-scale railway networks.

5.2.3. Recommendations from the PoC

Infrastructure Planning: Gaining insights from GraphSAGE's understanding of cause-effect relationships can significantly enhance infrastructure planning and network optimisation. In practical terms, the model's output can be applied as a basis for decision-making regarding infrastructure investments and improvements. For instance, if GraphSAGE always or frequently identifies certain nodes as recurrent sources of delays or incidents, this could imply that either the area of the operational network needs further optimisation or the infrastructure requires improvement. In addition, the model's understanding of the ripple effects of incidents can inform contingency planning, helping to create robust plans that account for potential downstream effects of disruptions. To capitalise on this, railway authorities should not only incorporate the insights generated by GraphSAGE into their planning processes but also value the outputs from data-driven approaches in decision-making.

Real-Time Response System: The potential of GraphSAGE to enhance real-time incident response systems is significant. Its continuous analysis and prediction capabilities can provide insights with the timely prediction of potential impacts of incidents, allowing for rapid and informed response strategies. Efforts should be channelled towards integrating GraphSAGE into the existing railway incident management systems to maximise its potential. However, this integration is not just a technical task, but it also requires significant cross-disciplinary collaboration. Artificial intelligence specialists need to work closely with railway operations experts to understand the practical requirements and constraints of the railway system, to ensure that the model can provide actionable insights that can be feasibly implemented in real-world operations.

Collaboration with Domain Experts: The successful implementation of GraphSAGE necessitates active collaboration with experts across different fields. As we mentioned, AI and data science specialists can ensure that the model is built, trained, and optimised effectively, but their technical expertise needs to be complemented by the practical insights of railway operations experts. These domain experts understand the unique characteristics and challenges of the railway system, and their insights can help guide the development of the model to ensure it accurately represents real-world conditions. They can also assist in interpreting the model's output in the context of operational realities and implementing its recommendations effectively. In essence, this collaborative approach brings together the strengths of AI technology and practical domain knowledge, to ensure that the model can be effectively applied in practice and provide meaningful improvements to railway operations.

6. Recommendations and Innovation Needs

6.1. Model Interpretability and Explainability

The application of predictive models in railway Traffic Planning and Management (TPM) demands a high level of transparency and explainability. Railway stakeholders, ranging from operational managers to engineers to system developers, need to understand how the predictive models reach their decisions, to ensure that these decisions are sound and actionable. For instance, if a model predicts potential delays or disruptions, it is crucial to know the basis for this prediction to make informed decisions about mitigating the delay. In addition, understanding the decision-making process can contribute to better trust in the model's output and foster a more receptive environment for the implementation of AI technologies in railway TPM.

The current heavily implemented deep learning models, such as neural networks, have demonstrated their prowess in handling complex tasks and delivering accurate predictions, but they suffer from a lack of interpretability. In other words, it is hard to decipher the "meaning" or significance of individual neurons or cells within the network. Also, it is challenging to understand how the model identifies and uses patterns in the data to reach its predictions. This inherent complexity, often referred to as the 'black box' problem, limits the ability of stakeholders to understand and trust the model's predictions. There are several techniques that can be employed to improve the interpretability of deep learning models:

- Feature Importance Analysis: This involves quantifying the contribution of each input feature to the model's predictions. This can provide insights into which aspects of the data the model deems most important, and by extension, which factors are most influential in causing the predicted outcome.
- Attention Mechanisms: These are a relatively new development in deep learning that allow the model to 'focus' on certain parts of the input data when making its predictions. By examining the areas where the model pays the most attention, we can gain insights into its decision-making process.
- Rule Extraction: This involves creating a set of rules or guidelines that mimic the behaviour of the deep learning model. These rules, which are usually simpler and more understandable than the underlying model, can provide a rough idea of how the model makes its decisions.

On the other hand, how to better understand the generated decisions from a well-interpreted model is another significant aspect that needs to be taken care of, that is, people from industrial sectors can benefit from understandable insights towards the entire context of the problem:

- Understanding the decisions made by the model can be facilitated by providing more contextual information. This could include background information about the data, the intended goals of the model, and the constraints it was operating under. By understanding the data used to train three different "Primary Delay Predictors", and their operating constraints, stakeholders can interpret their predictions more accurately. For example, if the model predicts a significant delay due to a snowstorm, but the stakeholders know that the snow removal equipment and personnel are already prepared (a piece of context the model might not be aware of), they can understand that the predicted delay might not be as severe as the model suggests.

- Visual tools like conceptual graphs, dynamic charts, or causal diagrams can help to explain the model's predictions in a more intuitive and understandable way. By visually depicting the relationships and dependencies in the data, these tools can help stakeholders grasp the basis of the model's predictions. For example, in the first stage of the second PoC, nodes and links are incorporated into a directed graph. Each node, visualised as a distinct point on the graph, represents an event or incident within the railway network, such as train delays or technical failures. Links, on the other hand, symbolise the cause-effect relationships between these events. For instance, a link might be drawn from a "Heavy Rainfall" node to a "Signal Failure" node, indicating that adverse weather conditions can cause a technical failure in the signalling system. The resulting directed graph provides a visual representation of the interplay of events within the railway network. By inspecting this graph, stakeholders can understand how different incidents are interconnected and anticipate the potential ripple effects that a single event can have, thus enhancing their comprehension of the GraphSAGE model's operation and outcomes.
- Natural language explanation is another effective way for translating the model's predictions and the reasoning behind them into natural language. This can make the model's output more accessible and understandable, especially for stakeholders who lack technical expertise in AI or data analysis.

6.2. Gather Comprehensive and High-Quality Data

As the fundamental part of any form of analysis, especially in complex environments such as railway operations, the process of gathering comprehensive and high-quality data cannot be oversighted. The data collection process is a critical phase that could determine the accuracy and reliability of the subsequent analysis, as the nature, scope, and quality of the data will directly influence the insights that can be derived. When dealing with railway operations, data gathering extends beyond mere collection to a meticulous selection of multiple sources. The scope of data to be collected is extensive, considering the myriad elements that comprise the railway system.

For instance, key among the data are (1) historical traffic patterns that offer insights into habitual railway usage, peak periods, and patterns over time. This data type can provide critical background information on recurring events and seasonal variations that significantly impact railway operations. The acquisition of (2) train schedules can reveal the operational mechanisms and patterns of the train system. It holds valuable data on train frequencies, route details, inter-station timings, and other scheduling intricacies that define the operation of the railway system. This data, when combined with the (3) real-time operation data, can offer insights into scheduling efficiency and areas of potential improvement. (4) Infrastructure details form another crucial component of the data gathering process. This encompasses data about the physical attributes of the railway system, such as the number of tracks, station details, condition of railway lines, maintenance schedules, and so on. Such data is vital in understanding the capacity, potential bottlenecks, and physical constraints of the railway system. (5) Incident records offer an invaluable resource for understanding the vulnerabilities and risks within railway operations. These records can provide a history of disruptions, technical failures, accidents, and other incidents that have impacted railway operations. Analysing such data can assist in identifying patterns, recurrent issues, and potential areas that need attention for safety and operational efficiency.

The process of data gathering also requires an effective system for storing and organising the data for easy access and analysis. It is essential to ensure data integrity, secure storage, and consistent updates for the most accurate and up-to-date insights. Gathering comprehensive and high-quality data is a dynamic process that requires consistent updates and revisions. Given the dynamic nature of railway operations, new data must be continuously gathered and integrated into the existing dataset. This iterative process helps keep the analysis relevant and adaptive to changing operational realities.

In conclusion, the data gathering process in railway operations analysis is a detailed and multi-faceted procedure. It is a task that demands a high level of diligence, precision, and a broad understanding of the railway system's operations. However, the rewards are significant, as comprehensive and high-quality data forms the foundation for effective and insightful railway operations analysis.

6.3. Choose Appropriate AI/ML or Hybrid Techniques

One of the significant goals when implementing the PoC in line with the requirements of this deliverable is to generally assess which AI/ML technique is potentially more suitable for addressing the challenges of a certain TPM problem. Based on the previous SWOT analysis for both PoCs, we summarised the highlights of several keynote AI techniques with their corresponding promising application areas as below:

Graph Neural Networks (GNNs): Consider GNNs when dealing with graph-structured data in railway networks. GNNs are powerful for modelling spatial dependencies, capturing network structure, and incorporating node and edge features.

Time Series Analysis (Spatial-Temporal NNs): If the problem involves forecasting or analysing temporal patterns, time series analysis techniques like ARIMA, LSTM, or Prophet may be suitable.

Reinforcement Learning (RL): RL can be applied when the problem involves sequential decision-making in dynamic railway environments, such as train scheduling or resource allocation. RL techniques like Q-learning, policy gradients, or deep RL can be explored.

Optimization Algorithms (Evolutionary Computing based): Especially outperformed on problems that involve optimising resource allocation, train scheduling, or route planning, optimization algorithms such as linear programming, integer programming, or genetic algorithms can be considered.

In summary, we recommend addressing a certain type of problem with specific AI techniques, it is hard to say some method is performing always better than others. The expected accuracy largely depends on the context of question.

Incorporating AI-based techniques with non-AI methods in a hybrid model is a promising approach in advancing the capabilities of predictive and analytical models, often leading to improved performance and more robust solutions. The synergy between these two distinct methodologies brings about a multi-dimensional perspective to problem-solving. AI-based techniques, with their sophisticated ability to learn complex patterns and make predictions from large datasets, can unearth insights that might be invisible to traditional non-AI methods. However, these techniques sometimes suffer from opacity and overfitting issues. In contrast, non-AI methods, while possibly lacking in their capacity to handle vast, complex datasets, often provide transparency and easy-to-interpret outcomes based on explicit, understandable rules or statistical techniques. Their strengths can, to some extent, counterbalance the limitations of AI techniques. In a hybrid model, the attributes of both approaches are leveraged, such that the strength of one compensates for the weakness of the other, leading to a more robust and versatile solution. This symbiotic relationship provides the opportunity to harness the full potential of both methodologies, enhancing the robustness of the model and its adaptability to diverse problem domains.

Ensemble Methods: Explore ensemble methods that combine the predictions or outputs of different AI-based/ML models and non-AI methods. This can lead to improved accuracy, robustness, and generalisation. Techniques such as model averaging, stacking, or boosting can be employed to create an ensemble of diverse models, each utilising different AI-based/ML techniques or non-AI methods.

Rule-based Systems: Combine AI-based/ML models with rule-based systems or expert knowledge to incorporate domain-specific rules, constraints, or heuristics. Rule-based systems can provide interpretability, explainability, and the ability to incorporate human expertise into the decision-making process. AI-based/ML models can then learn from data and optimise the decision process within the boundaries set by the rule-based systems.

Performance Monitoring and Feedback: Continuously monitor the performance of the hybrid methods and gather feedback from domain experts and stakeholders. This iterative feedback loop helps identify any limitations or issues in the hybrid approach, refine the models, update rules or constraints, and enhance the overall performance.

6.4. Ensure Scalability and Real-Time Capabilities

As a complex railway system generates vast amounts of data in a continuous time manner, it is imperative to use appropriate data storage, processing, and querying techniques. These are the bedrock of a well-structured data pipeline that ensures data integrity, accessibility, and efficient processing. The choice of technology should factor in the data volume, velocity, and variety, typical of railway system operations.

Data Storage and Processing: Distributed computing frameworks such as Apache Hadoop and Apache Spark have been designed to distribute computation burdens across multiple nodes, enabling large-scale data processing. Hadoop's distributed file system enables data storage across multiple machines, while Spark provides fast, in-memory processing, which is particularly useful for iterative machine learning tasks.

Data Querying and Retrieval: This becomes particularly effective as the size of the data grows. Techniques like indexing, partitioning, and optimization of query languages can enhance the speed of data retrieval and ensure that analytical tasks and reporting do not become bottlenecks.

Parallel Processing Techniques: The use of parallel processing techniques can significantly speed up computation efficiency. By dividing tasks and executing them simultaneously across multiple cores or processors, parallel processing can handle complex computations and large datasets more rapidly.

In addition to the above, several suggestions can be put forth to enhance the real-time capabilities of railway operations.

Low-latency Architectures: Designing the system with low-latency architectures enables quick response times, which are critical for real-time decision-making. This might involve reducing processing delays, optimizing algorithms for speed and efficiency, and ensuring efficient data retrieval. Techniques such as load balancing, efficient use of caching, and streamlined data pipelines can help in achieving low-latency performance.

Edge Computing: Given the geographical spread of railway networks, exploring edge computing approaches can be beneficial. Edge computing brings computation closer to the data source, reducing the time taken to send data to a central location for processing. This localised processing can improve response times and make real-time decision-making more viable.

Online Learning Techniques: As railway operations continue to generate data in real-time, online learning techniques that allow the model to continuously learn and adapt to new data can be beneficial. Online learning can ensure that the model's performance does not degrade over time and that it stays responsive to the most recent data trends.

By considering and incorporating these recommendations, vast data generated by railway systems can be properly accommodated and hence real-time, data-driven decision-making can be facilitated.

6.5. Collaborate with Domain Experts

Effective collaboration with domain experts, such as railway operators, traffic planners, and incident managers, is crucial for the success of AI-based solutions in railway systems. Their expertise and insights can ensure the relevance and practicality of these solutions, infusing them with a deep understanding of the intricacies of the railway system.

Human-in-the-Loop: Incorporating human expertise into the machine learning process, known as the "Human-in-the-Loop" approach, is invaluable. By doing so, we can identify the key factors that influence the traffic planning process and ensure that our AI-based models consider these factors. The domain experts can help verify and validate that the model aligns with the practical constraints and requirements of the system design policies, which in turn enhances the model's applicability and efficacy. Moving on to the deployment phase, the input of domain experts is crucial in validating the effectiveness of the AI solution in real-world scenarios.

Engagement in the Deployment Phase: By involving domain experts in the deployment phase, we can test and validate the AI solution under real-world conditions. These experts can provide valuable feedback on the system's performance, suggest improvements, and help fine-tune the model to better suit the real-world operational context. Consideration of potential biases in data, as well as safety, privacy, and regulatory aspects, is necessary to ensure the responsible and ethical application of AI-based models.

Consideration of Potential Biases: Discussing and acknowledging potential biases in the data collection and processing stages is crucial. Left unchecked, these biases can distort the model's predictions and lead to unfair or inefficient outcomes. Collaborating with domain experts can help identify and mitigate these biases.

Safety, Privacy, and Regulatory Considerations: Pure AI-based models may not fully consider these issues during experimental design. Therefore, it is important to engage domain experts to ensure that the model's deployment aligns with safety standards, respects privacy norms, and complies with relevant regulations.

To conclude, involving domain experts throughout the development and deployment of AI-based solutions can greatly enhance their practicality, effectiveness, and acceptance among end-users.

7. Conclusions

This document reported the identification of possible future innovation needs and recommendations in the railway industry to enhance the efficiency of railway traffic planning and management process. It addressed a detailed analysis of the two proofs-of-concept (PoCs) proposed in the previous WP4 deliverables, namely, “Graph Embedding based Primary Delay Prediction” and “Big Data on Incident Attribution Analysis”.

The report critically examined the outcomes of the PoCs, discussing recent advancements in their respective fields, and conducting SWOT analyses to identify the main Strengths, Weaknesses, Opportunities, and Threats. Some recommendations emerged that aimed to address the identified challenges in order to enhance the technical and operational feasibility of the proposed AI approaches. Specifically, as for ”Graph Embedding based Primary Delay Prediction”, recommendations are mainly oriented at: i) introduce relevant AI or non-AI techniques that enhance interpretability; ii) incorporate additional features into the learning process; iii) emphasise the ability of Customization and Adaptation of the model. On the other hand, regarding “Big Data on Incident Attribution Analysis”, recommendations encompass: i) explore the feasibility of enhancing infrastructure planning and network optimization; ii) integrate the current framework into a Real-Time Response System; and iii) Collaborate with Domain Experts.

Eventually, general recommendations and innovation needs for future developments in railway planning and management tasks have been given. These include i) Improving Model Interpretability and Explainability; ii) Gathering comprehensive and high-quality data; iii) Choosing appropriate AI/ML or hybrid techniques; iv) Ensuring scalability and real-time capabilities; v) Collaborating with domain experts. All the recommendations presented in this document will converge into the definition of roadmaps of AI in railways which will be discussed in our next Deliverable 5.3.

Bibliography

[1] RAILS, “Deliverable D1.1: Definition of a reference taxonomy of AI in railways,” October 2020. [Online]. Available: https://rails-project.eu/downloads/deliverables/

[2] N. Bešinović, L. De Donato, F. Flammini, R. M. P. Goverde, Z. Lin, R. Liu, S. Marrone, R. Nardone, T. Tang, and V. Vittorini, “Artificial intelligence in railway transport: Taxonomy, regulations, and applications,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 14 011–14 024, 2022.

[3] RAILS, “Deliverable D1.2: Summary of existing relevant projects and state-of-the-art of AI application in railways,” 2021. [Online]. Available: https://rails-project.eu/downloads/deliverables/

[4] R. Tang, L. De Donato, N. Besinović, F. Flammini, R. M. Goverde, Z. Lin, R. Liu, T. Tang, V. Vittorini, and Z. Wang, “A literature review of artificial intelligence applications in railway systems,” *Transportation Research Part C: Emerging Technologies*, vol. 140, p. 103679, 2022.

[5] M. J. Pappaterra, F. Flammini, V. Vittorini, and N. Bešinović, “A systematic review of artificial intelligence public datasets for railway applications,” *Infrastructures*, vol. 6, no. 10, p. 136, 2021.

[6] RAILS, “Deliverable 1.3: Application Areas,” 2021. [Online]. Available: https://rails-project.eu/downloads/deliverables/

[7] ——, “Deliverable 4.1 – WP4 Report on case studies and analysis of transferability from other sectors (Railway traffic planning and management),” 2022. [Online]. Available: https://rails-project.eu/downloads/deliverables/

[8] R. Tang, R. Liu, and Z. Lin, “Predicting primary delay of train services using graph-embedding based machine learning.” *SSRN-preprint*, 2023. [Online]. Available: http://dx.doi.org/10.2139/ssrn.4472063

[9] RAILS, “Deliverable 4.2 – WP4 Report on AI approaches and models,” 2022. [Online]. Available: https://rails-project.eu/downloads/deliverables/

[10] ——, “Deliverable 4.3 – WP4 Report on experimentation, analysis, and discussion of results,” 2023. [Online]. Available: https://rails-project.eu/downloads/deliverables/

[11] D. Wang, P. Cui, and W. Zhu, “Structural deep network embedding,” in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2016, pp. 1225–1234.

[12] Y. Bengio, “Learning deep architectures for AI,” *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[13] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,” *IEEE Signal processing magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[15] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proceedings of the 2013 conference on empirical methods in natural language processing*, 2013, pp. 1631–1642.

[16] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[17] J. R. Quinlan, "Simplifying decision trees," *International journal of man-machine studies*, vol. 27, no. 3, pp. 221–234, 1987.

[18] T. K. Ho, "Random decision forests," in *Proceedings of 3rd international conference on document analysis and recognition*, vol. 1. IEEE, 1995, pp. 278–282.

[19] L. V. Shavinina, *The international handbook on innovation*. Elsevier, 2003.

[20] D. Yang, Z. Ma, and A. Buja, "A sparse svd method for high-dimensional data," *arXiv preprint arXiv:1112.2433*, 2011.

[21] W. L. Hamilton, R. Ying, and J. Leskovec, "Representation learning on graphs: Methods and applications," *arXiv preprint arXiv:1709.05584*, 2017.