Horizon 2020
European Union Funding
for Research & Innovation

Graph-SAGE based Railway Incident Attribution and cause-effect relation prediction

Ruifan Tang
PhD Student in Institute for Transport Studies
University of Leeds

Railway Problem Description

Objective of the PoC

Investigate cause-effect relations between incidents and predict how a specific delay at a certain location affects the wider network

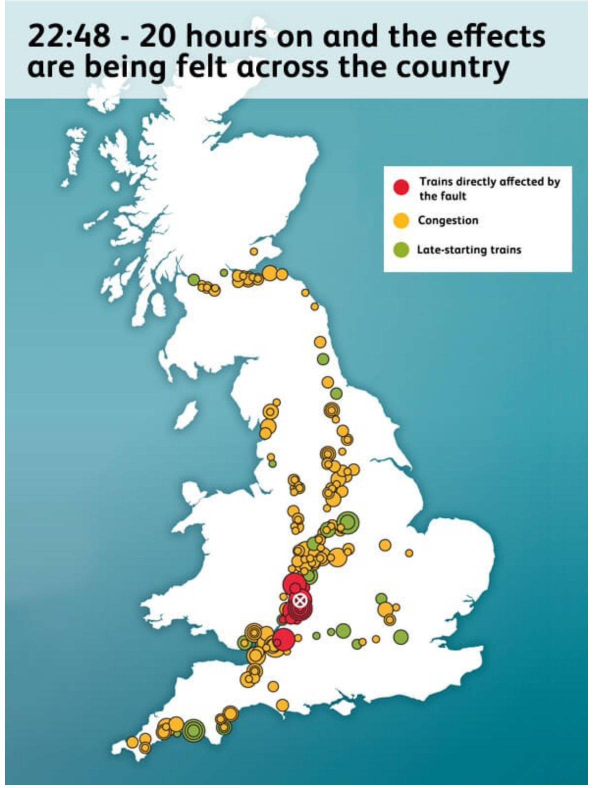
Constraints / Requirements

Understand complex non-linear spatio-temporal Interactions between incidents/trains
In-depth insights into delay propagation for better understanding of network-wide effects

Main Issues and Challenges

Limited and incomplete data

Generalization to diverse railway systems



Key Performance Indicators

Prediction Accuracy
Computational Time
Effectiveness

Proof-of-Concept as a Benchmark

AI Application
Machine Learning

AI Related Disciplines
Graph Neural Networks

AI Techniques
Semi-supervised: GraphSAGE
It utilizes labelled nodes during training
to generate representations for unknown nodes

Inspiring Solutions
Neighborhood Information Aggregation
Graph Representation Learning
Graph Visualization
Link Prediction

Datasets
Real-world Network data
Network Rail Open Data portal

Developments / Implementations
Capture the inherent structural patterns/relations
Performing a graph visualization module
Conduct a link prediction between labelled nodes

Exploited Software and Framework
Stellargraph, PyTorch Geometric, Sklearn, Keras

Hardware Requirements
Google Colaboratory
GPU(s) with CUDA cores and 16GB System RAM

Approach: A Modular Architecture

The proposed framework consist of:

- Big Data-based visualizations that incorporate the complex interactions between modelled train services and events.
- 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.

Each incident: event time, occurrence location, responsible train, incident reason, affected train or service group

Each service: departure/arrival time, passenger numbers, margin time, speed limit, rolling stock type

Network features: adjacency matrix, connectivity, link travel time, importance of stations

3D interactive visualization

Intervention simulation

GraphSAGE-based Incident/delay attribution

Performance improvement

Derive delay propagation pattern

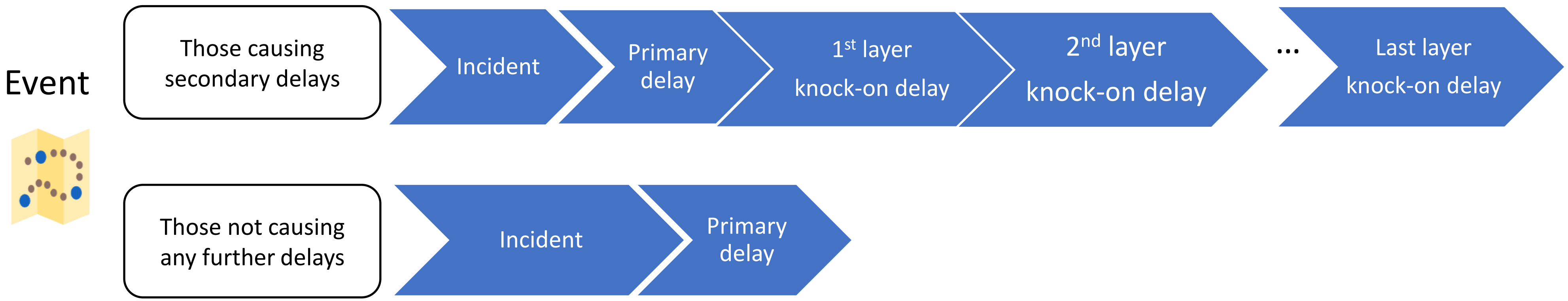
Description of dataset

- **Two main sources:**

The cooperating railway undertaking **TransPennine Express** provides the historic real-world operation timetables including train route information and their corresponding departure/arrival/dwell time, etc.

Delay attribution data collected from **Network Rail Open Data portal** including incident and delay durations, type of incidents, responsible/affected service(s), and reactionary delay timestamps/locations, etc.

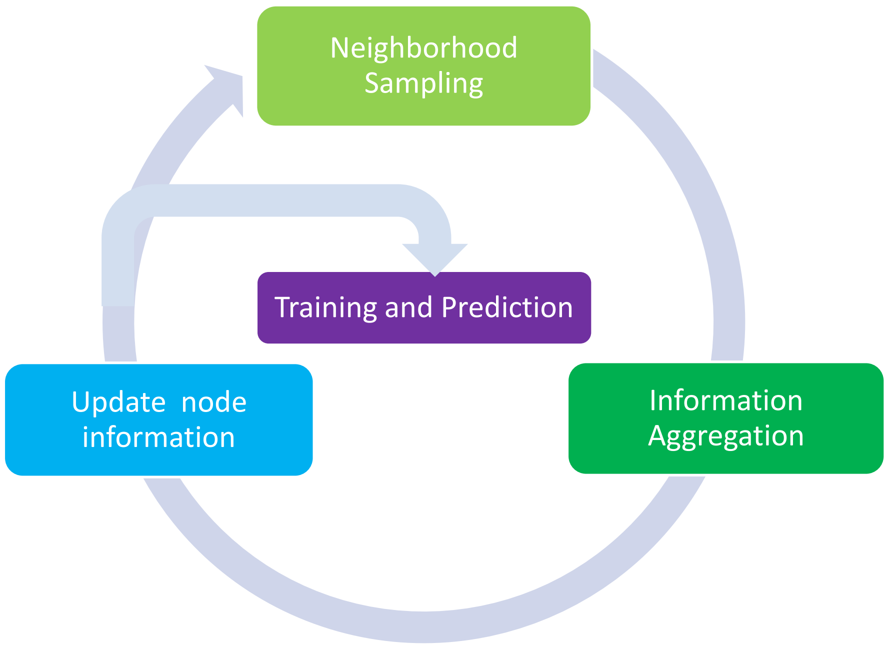
- **Data Explanation**



Module Implementation

- **3D interactive visualization**
  1) Extract all the events (i.e., individual incidents)
  2) Create a node list for all the train services in the network (those with primary/secondary delays or both), and combine it with the list generated in 1)
  3) Create primary delay reaction set & knock-on delay reaction set for the two-categorized events respectively, according to responsible/affected train in delay attribution dataset.

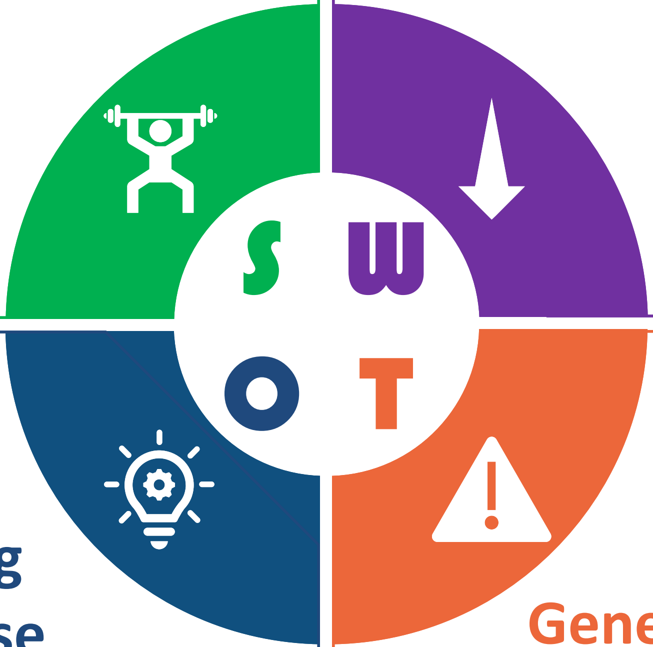
- **GraphSAGE-based model**
  1) Sampling and normalising all the attributes for nodes (i.e., individual incidents) an based on which create a graph.
  2) Define the distance (e.g., how many hops) between two nodes.
  3) Define the desired aggregation strategy (i.e., Mean/Max-Pooling/Attention-based Aggregation)
  4) Process the label attributes – classification/regression.



SWOT Analysis of the Investigated Approach

Scalability
Inclusion of Node Features
Inductive Learning
Flexibility

Data availability and quality
Interpretability



Proactive decision-making
Real-time incident response
Network optimization

Generalization to different contexts
Data scalability

Thank you for your attention!

RAILS Roadmaps for AI Integration in raiL Sector

- ✓ *Deliverable D4.1: WP4 Report on case studies and analysis of transferability from other sectors (Railway planning and management)*
- ✓ *Deliverable D4.2: WP4 Report on AI approaches and models*
- *Deliverable D4.3: WP4 Report on experimentation, analysis, and discussion of results*
- *Deliverable D4.4: WP4 Report on identification of future innovation needs and recommendations for improvements*

Available at: https://rails-project.eu/downloads/deliverables/