Cross City Urban Transit Networks Analysis

Group 16 : Ravinithesh, Nitish
Network Machine Learning Course Project - Spring 2023

Aims

- How do the topological properties of public transport networks differ across different cities and compare in terms of connectivity?
- Which public transport stops are the most important in each city based on their centrality measures?
- Whether node features can predict directed edges between them?

Exploration

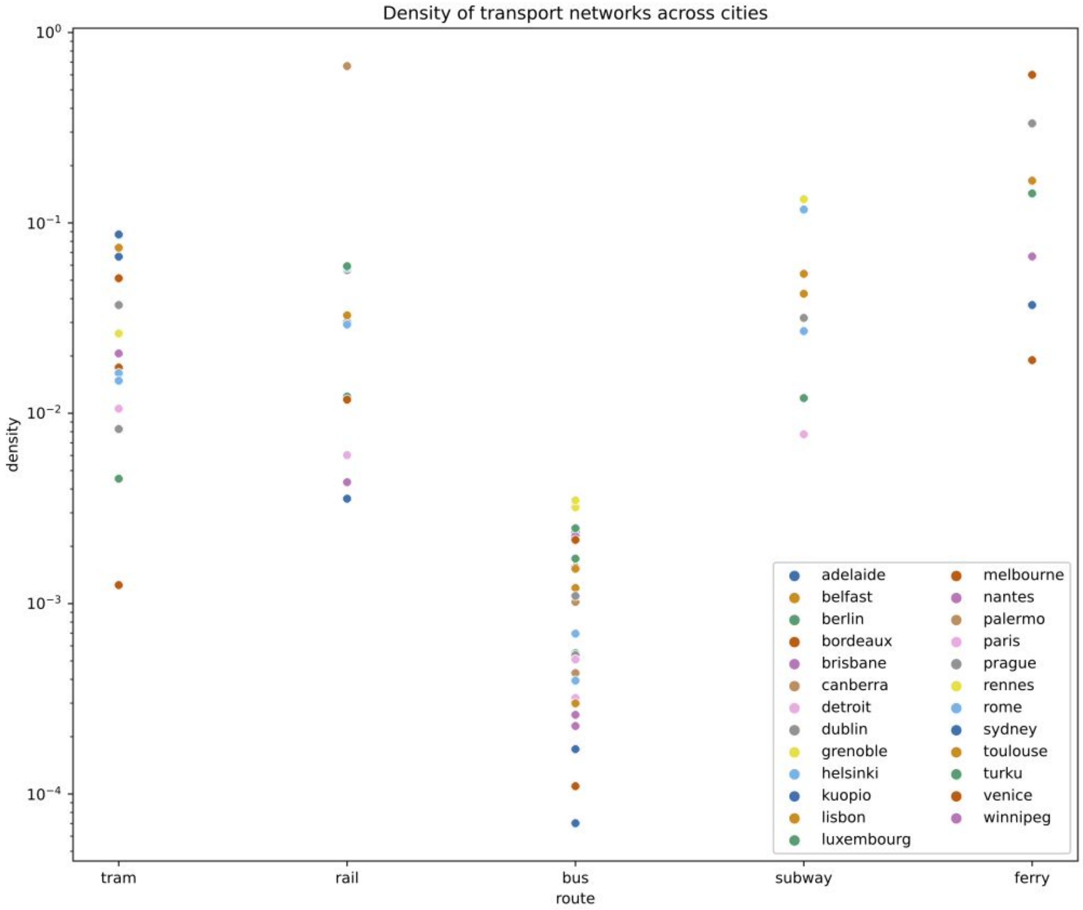
Preprocessing

- Two Main files for each city
    - Nodes File: Node Id, Name, Latitude and Longitude
    - Edges File
        - From and To nodes
        - Straight line distance
        - Average duration between stops
        - Flow data (per hour)
        - Type of transport
    - Representation: nx.MultiDiGraph
- Merged nodes with same name
- Removed self loops between the nodes
- Used directed graphs throughout the analysis

Id	Name	Id	Name
1001	Lausanne Gare	1001	Lausanne Gare
1004	Lausanne Gare		

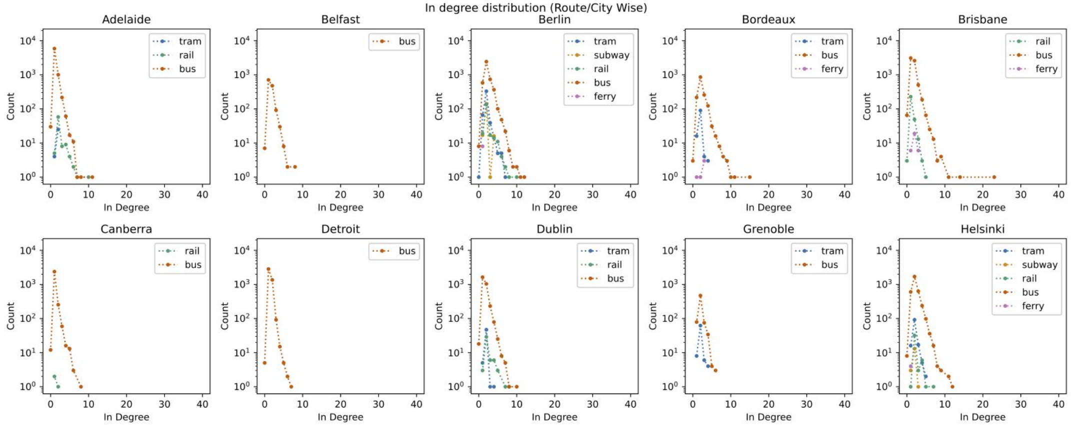
Modes of Transport

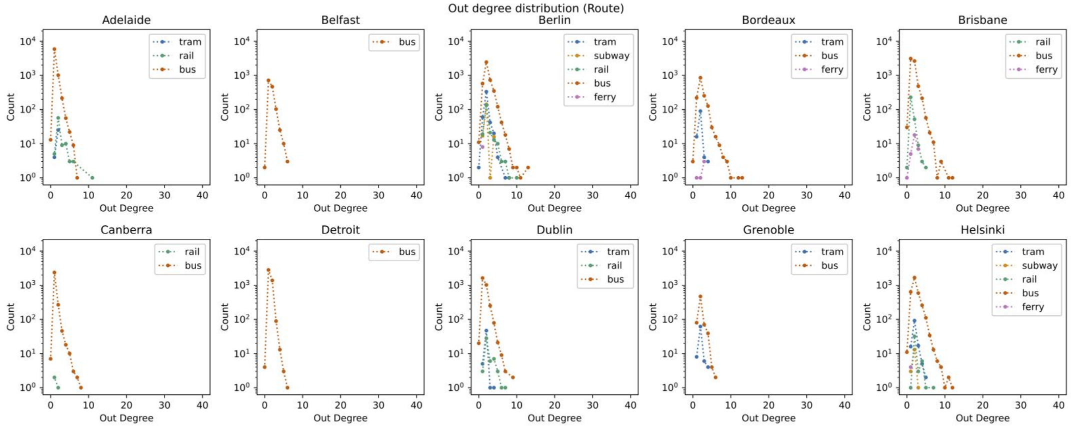
- There are 7 possible route types, none of the cities in the dataset have furniculars / gondolas
- Highly dense ferry networks
- Low density of Bus networks.
- Subway, Tram and Rail networks lie between them.



Degree Distribution

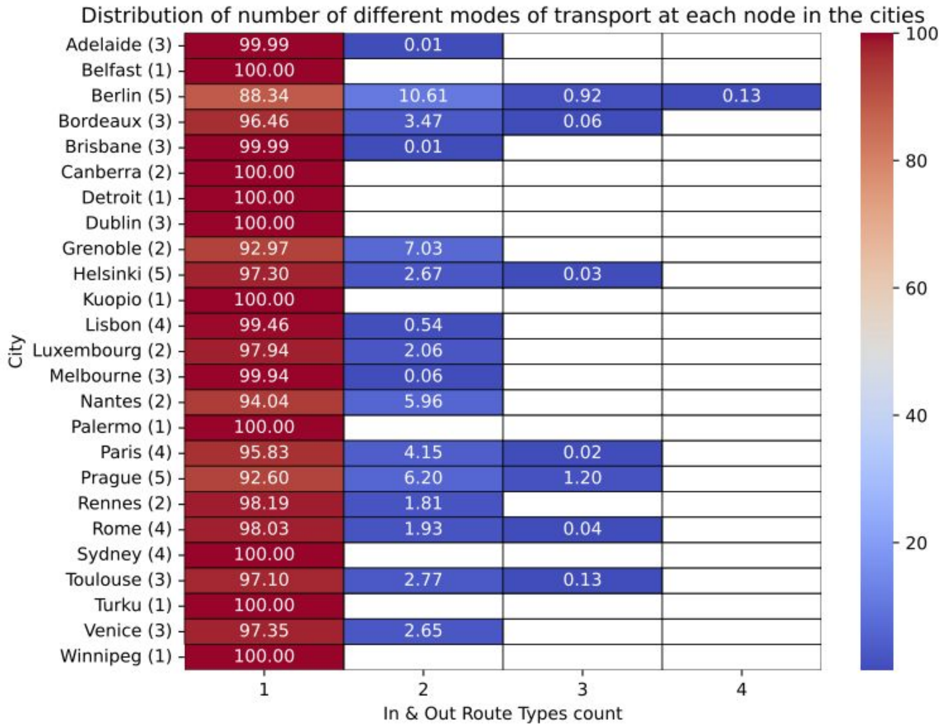
- Many nodes with low degrees and few nodes with higher degrees.
  - Cannot be modelled by a random network model.
- Presence of nodes with 0 in/out degrees
- The average degree has a median of 2.03 and mean of 2.0 telling us that each station is connected with two other stations.





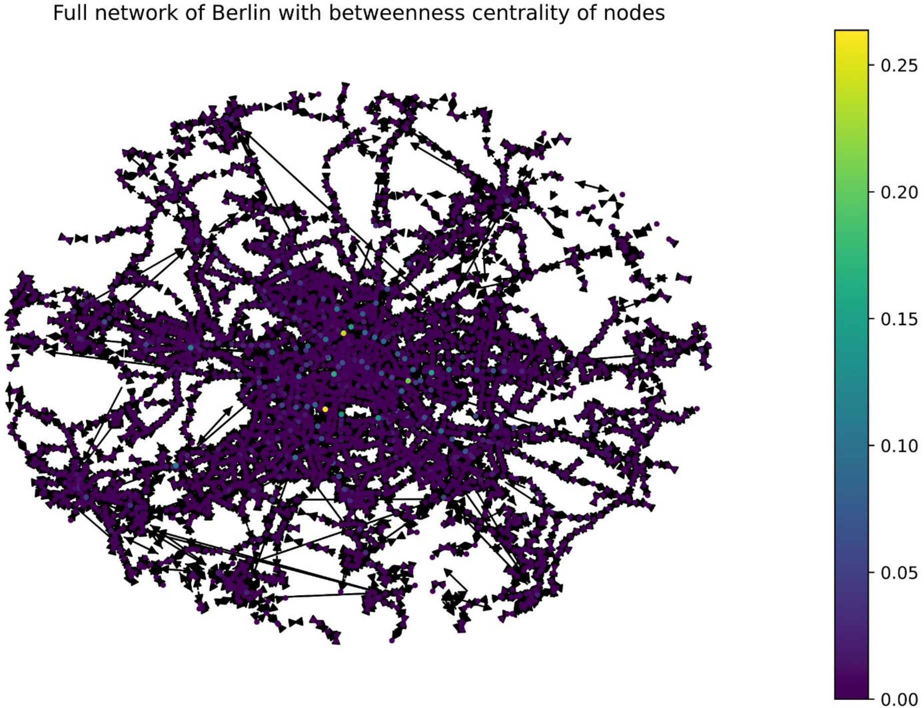
Interchangeability

- Changing between modes of transport
- Skewed - Cities which have nodes majorly connected with few (typically 1) mode of transport.
- Diverse - Cities which have a good proportion of nodes connected to each other by more than 1 mode of transport



Centrality Measures

- Central/Transit stations have high betweenness centrality
- In Berlin, they are the transit stations between different lines of the S-Bahn, regional and long-distance trains.

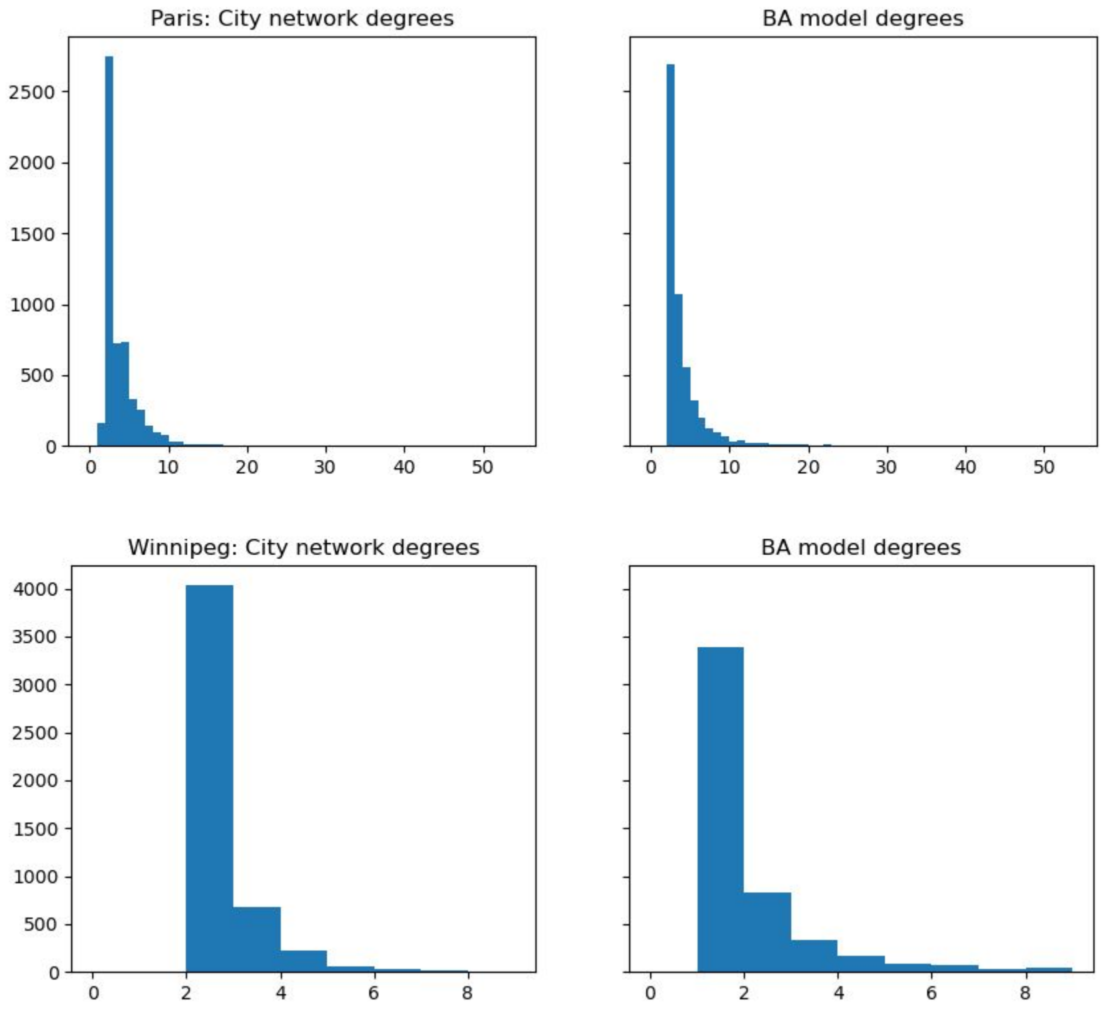


Clustering Coefficients

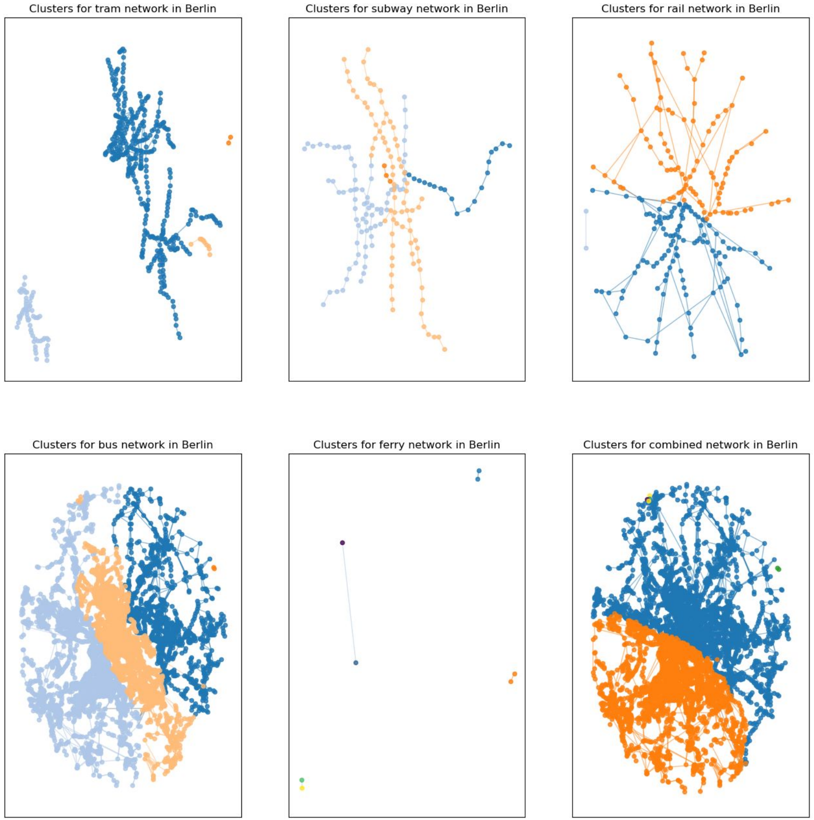
- Average Clustering Coefficient (ACC) can provide insight into the local connectivity of the network
- Almost all modes have low clustering coefficients
  - Berlin ferry has 0 some even zero, indicating disconnected components like the The rail and bus modes have higher ACC, suggesting that they are better connected than the other modes of transport. Cities have one (or rarely two) modes with a high clustering coefficient.

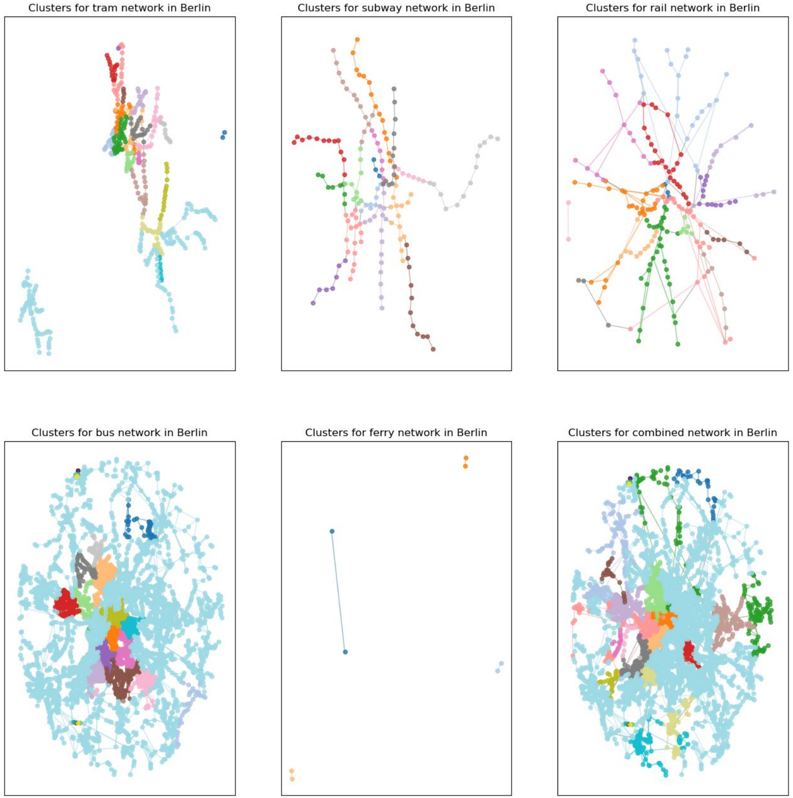
Task - Modelling networks

- Looking at the degree distribution above, we concluded that it cannot be modelled by a random network model.
- Comparing it with the Barabasi Albert generation model with an appropriate choice of q (from the degree distribution of network nodes).



Spectral clustering & Community detection

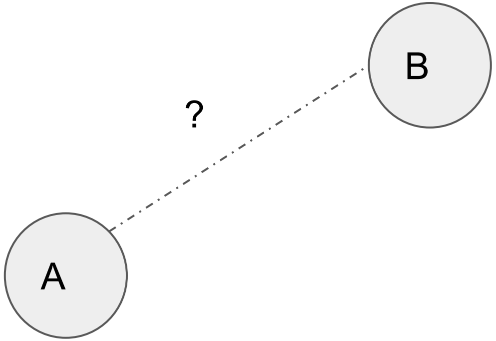


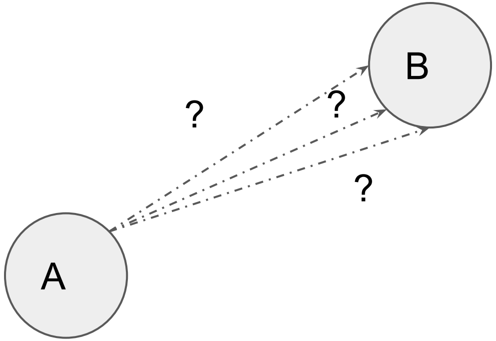


Exploitation

Prediction Tasks

- Directed Edge Prediction between nodes in individual transport mode for each city.
- Edge Label prediction between nodes in a complete network for each city.



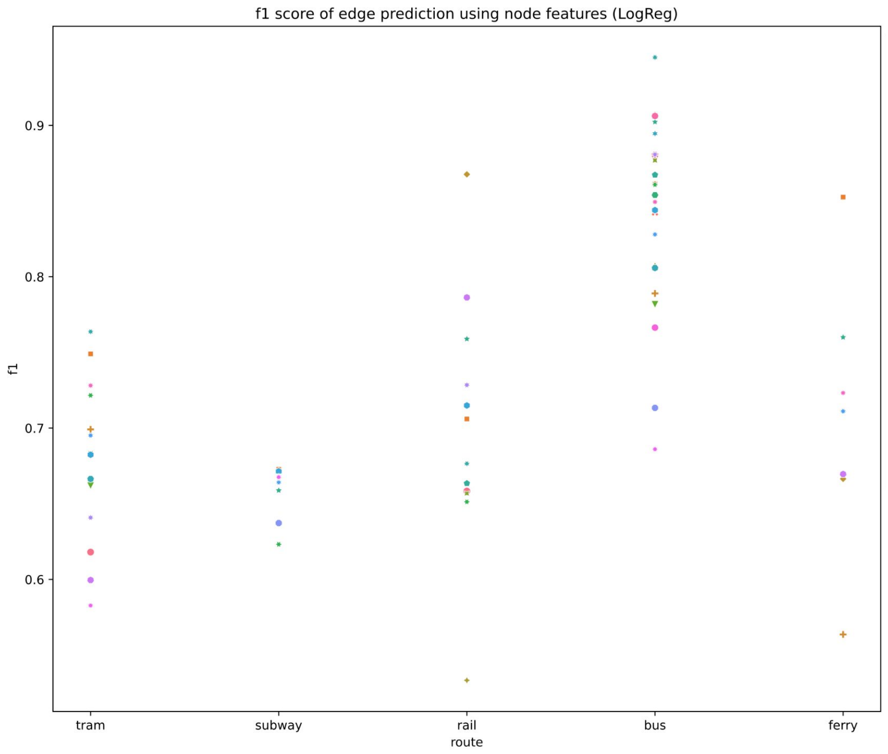


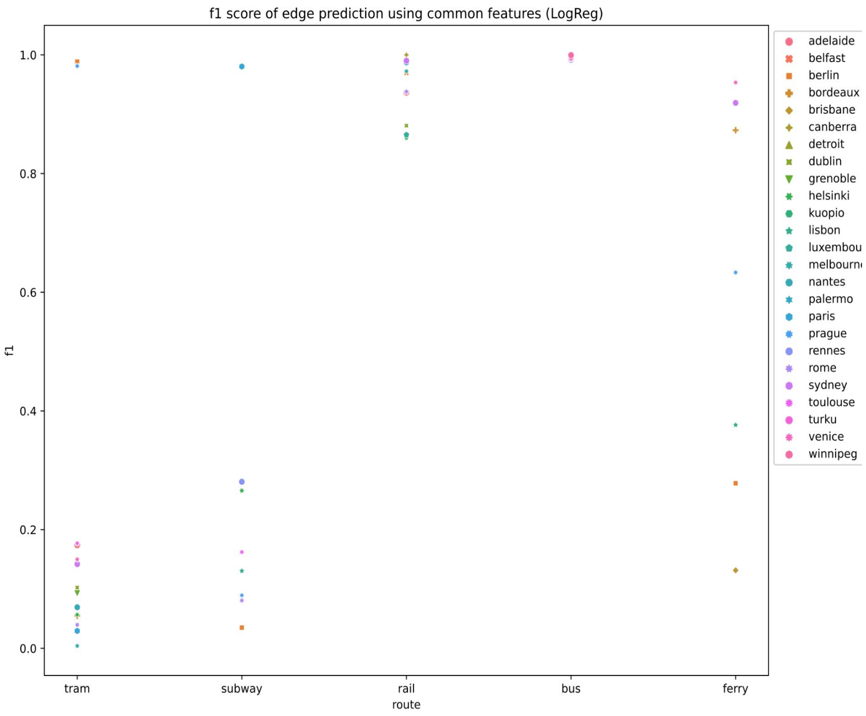
Features & Setup

Handcrafted features: (node)	Handcrafted features: (common)	Node2vec features
1. In-degree centrality 2. Out-degree centrality 3. Betweenness centrality 4. Katz centrality	1. fraction of the number of shared incoming neighbours 2. fraction of the number of shared outgoing neighbours	• Capture the structural properties of nodes in the neighbourhood.

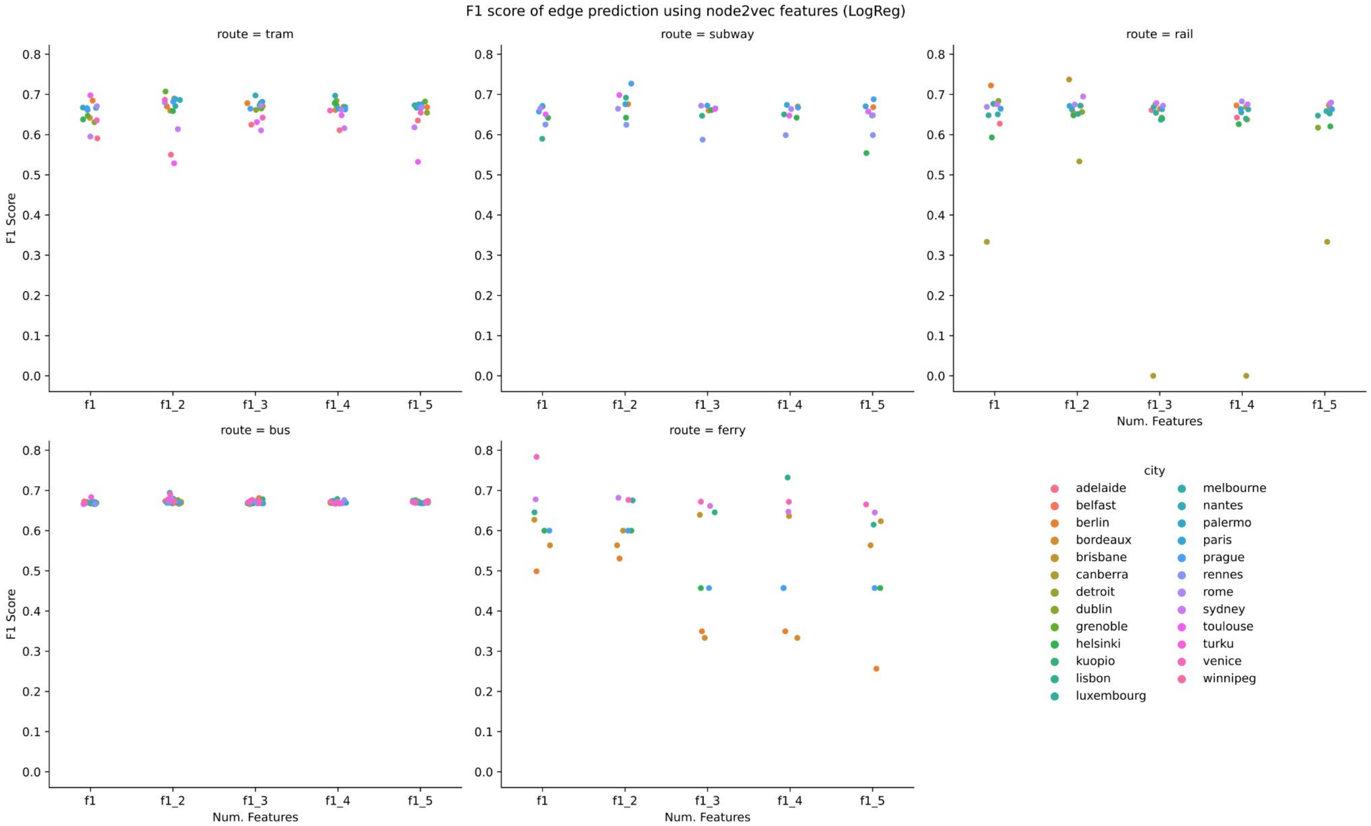
- Difference in node features are taken
- 70 - 30 train / test split
- Logistic Regression

Edge Prediction - 1

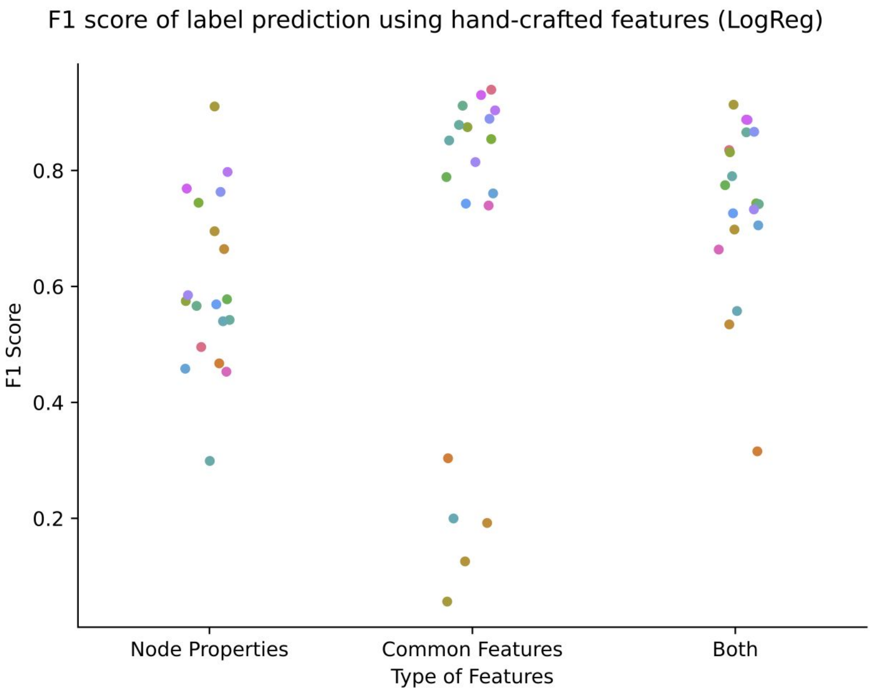


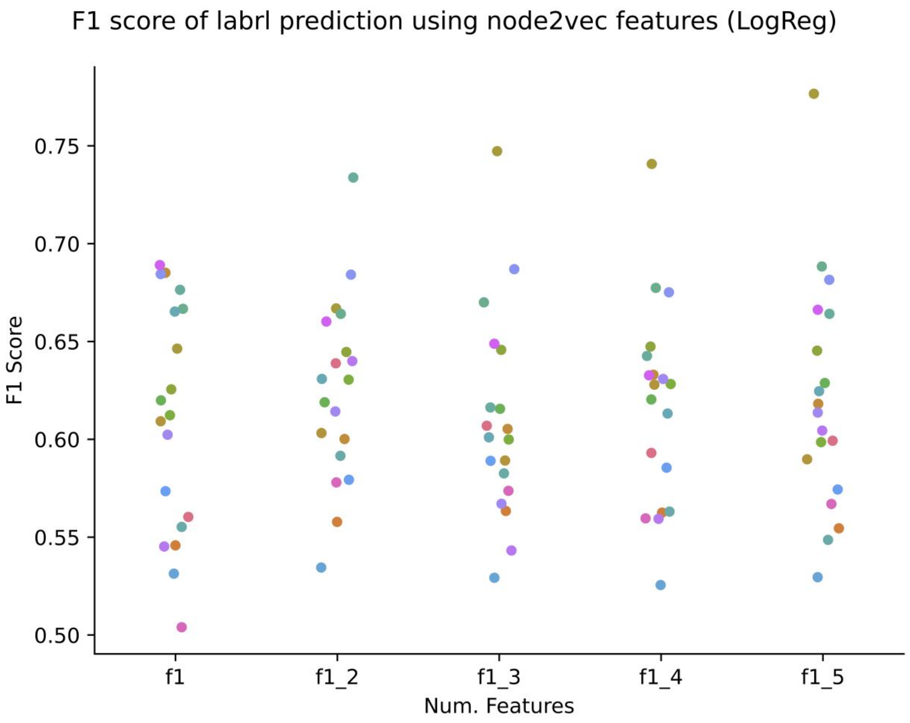


Edge Prediction - 2



Label Prediction





Graph Neural Networks

- No consistent performance with the hand-crafted and Node2Vec features across all cities.
- Experimented with GNNs.
- The results did not have satisfactory performance.
- Reasons
    - Complexity and inherent noise within transport networks
    - Limited amount of labelled training data
    - Imbalance in the examples for different modes of transport.

Summary

- Analysed 25 cities transport networks
- Low density and clustering coefficient of the networks and existence of hubs
- Handcrafted node properties were suited for edge prediction.
- Node and common properties together were suited for label prediction
- Node2Vec features were good only in few instances.
- Geographical dimensions of the city and other factors, like its population, area, and usage of public transport, were not part of the analysis.

Thank you