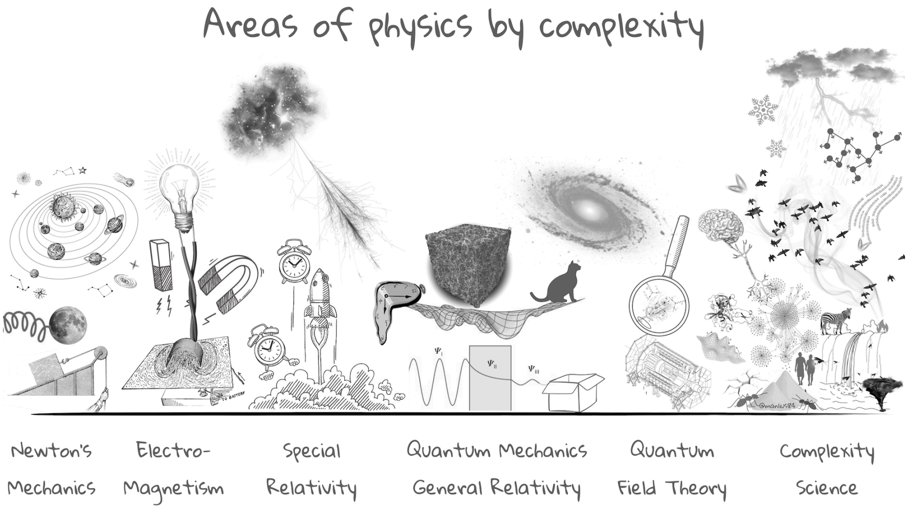
Università degli Studi di Padova



Pietro Scapolo's project

Date of examination: 25/06/2024.

Last update: June 24, 2024

Final Report

Physics of Complex Networks: Structure and Dynamics

Contents

1 | #42 Public Transport in the UK

1.1 | Task Explanation

The objective is to create a network for UK public transport systems, focusing on different types of transport available as CSV datasets. The goal is to generate individual networks for each city, excluding less populated ones. For each city and each layer, two files have been produced: one for nodes (nodeID, latitude, longitude, geometry data) and one for edges (nodeID_from, nodeID_to). The challenge lies in integrating two datasets that do not directly correspond with each other.

1.2 | Methodology

Population Data Redistribution

Due to difficulties in matching nodes to cities using provided datasets, I used counties and travel-to-work areas (TTWAs) instead. The population data were initially at the county level. For a more accurate representation of transport networks, I redistributed the population data using the ‘st_intersection‘ function from the ‘sf‘ library, assigning the population to each TTWA proportionally based on the intersection area with the counties. Population data for counties was sourced from the ONS [2]. Mathematically, if $A_{i,j}$ is the area of intersection between county i and TTWA j, and P_i is the population of county i, the population assigned to TTWA j from county i is:

$$P_{i,j} = P_i \times \frac{A_{i,j}}{A_i} \tag{1.1}$$

where A_i is the total area of county i. The total population for TTWA j is then:

$$P_j = \sum_i P_{i,j} \tag{1.2}$$

Precision and Recall Test

To verify the accuracy of this redistribution, I conducted precision and recall tests by comparing the population in TTWAs with the population data for each city sourced from the World Population Review [1]. Precision is calculated as:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1.3}$$

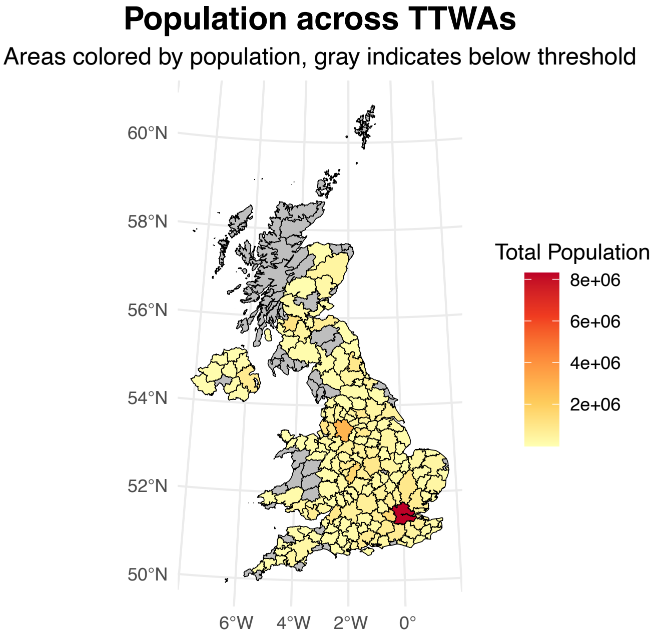
Recall is calculated as:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{1.4}$$

The results showed precision of 96.5% and recall of 97.2%, indicating the method successfully preserved the total population.

Filtering TTWAs by Population

To focus on significant areas, I excluded TTWAs in the lowest quartile of the population distribution. This threshold is adjustable based on analysis requirements.



Creation of Node and Edge Files

I created node and edge files for each TTWA. Nodes were assigned to TTWAs using their geographical coordinates with the ‘st_within‘ function from the ‘sf‘ library. Each node file includes nodeID, latitude, longitude, layer, degree, and geometry data for plotting. Edges were attributed to TTWAs based on the nodes they connect. Each edge file includes nodeID_from, nodeID_to. This resulted in node and edge files for each TTWA and the entire UK network, enabling detailed analysis at both local and national levels. TTWA boundary data was obtained from the ONS [3].

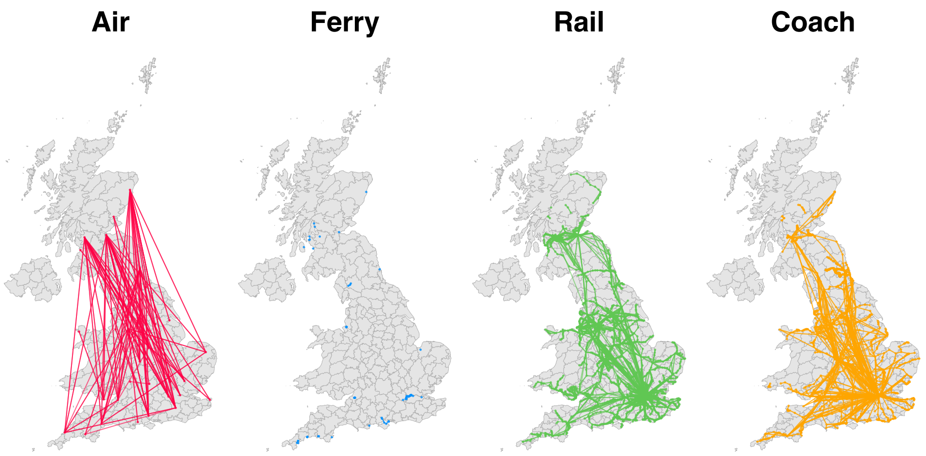


Figure 1.1: Visualization of the UK transport network across multiple layers, highlighting only the most populated regions.

Figure 1.2: Visualization of London transport network across multiple layers.

2 | Appendix

2.1 | #42 Public Transport in the UK

The reason I excluded transport lines and year data

The decision to exclude transport lines and year data was based on several considerations. The main objective of this study was to analyze the structural properties of the UK transport networks, and including detailed line information and historical data would have added unnecessary complexity.

The provided datasets were limited to one week of data from 2011, as detailed by Gallotti et al. (2015) [**?**], making it impossible to include multiple years. The datasets contained timetable information for every service departure, resulting in a voluminous and repetitive data set, complicating analysis. While Gallotti et al. used this data to study temporal network evolution, our focus was on network structure.

Handling the temporal data file, events.txt, was challenging due to its size. The only potential way to implement this would have been to reconstruct the lines from the events.txt file, which would have been extremely complex. One approach could have been to identify departure and arrival nodes and use the provided stop times to determine which nodes formed a line but that would have required advanced algorithms or machine learning techniques.

Given these challenges, I decided to focus on the more feasible task of network structure analysis, leaving the detailed reconstruction of transport lines for future research. This approach allowed for a clear and manageable analysis of the network's structural properties

Additional Plots: Counties vs. TTWAs Visualization

To provide a clearer understanding of how the population data were redistributed from counties to TTWAs, the following plots illustrate the geographical relationship between these regions.

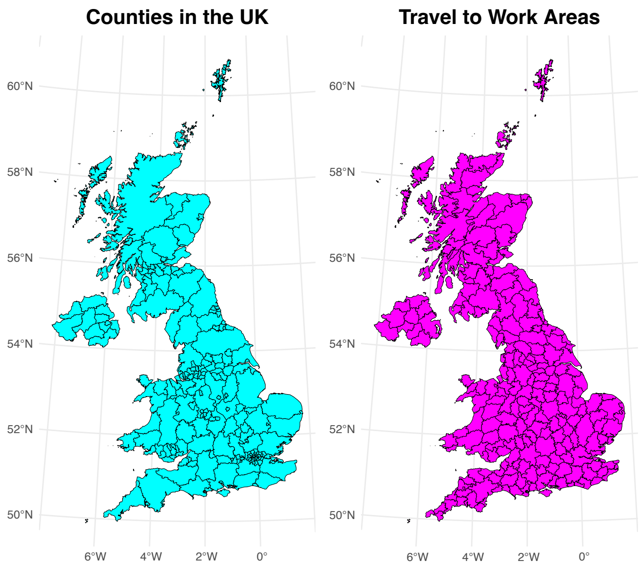


Figure 2.1: Visualization of counties and TTWAs.

Further Analysis of the Interlayer Network

Further analysis of the interlayer network could provide insights into the robustness and efficiency of the UK's transport systems. This section outlines potential avenues for such analysis:

Attack Methods on Network Robustness

Studying how different types of attacks (e.g., random failures, targeted attacks) affect the network's connectivity can reveal vulnerabilities. For example, simulating the removal of nodes or edges can help identify critical components whose failure would significantly disrupt the network.

Shortest Path Analysis Using Random Walkers

Random walker algorithms can be employed to estimate the shortest paths within and across different transport layers. This method can highlight the efficiency of the network in terms of travel time and connectivity, providing practical insights for optimizing route planning and improving transport services.

Other Advanced Methods Studied During the Course

Other methods such as centrality measures, community detection, and flow analysis could be applied to further dissect the network's characteristics. For instance:

- **Centrality Measures:** Identifying the most influential nodes in the network based on degree centrality, betweenness centrality, or eigenvector centrality.
- **Community Detection:** Using algorithms like Louvain or Girvan-Newman to find clusters or communities within the network, which could correspond to functional regions or densely interconnected areas.
- **Flow Analysis:** Analyzing the flow of passengers or goods through the network to understand bottlenecks and optimize network performance.

These analyses can provide a deeper understanding of the transport network's dynamics and offer data-driven recommendations for enhancing its resilience and efficiency.

3 | Bibliography

[1] Population of Cities in the UK. https://worldpopulationreview.com/countries/cities/united-kingdom. [Accessed 29-Mar-2023].

[2] County Population Data. https://geoportal.statistics.gov.uk/datasets/4832bb99c4844971b111739be44f1bc5_0/explore?location=52.747048%2C0.003210%2C7.76&showTable=true. [Accessed 29-Mar-2023].

[3] Travel-to-Work Areas Boundary Data. https://geoportal.statistics.gov.uk/datasets/7e6f5db8adec4bb984a99a2d91fcf3f4_0/explore. [Accessed 29-Mar-2023].