Module:	Data Analytics and Visualization
Academic Year:	2023-24
Lecturers:	Prof. W. Guo, Dr. I. Petrunin
Title:	**Graph Analysis of Air Transport Network**
Date issued:	November/December 2022
Submission date:	*To be notified on Canvas*.
Estimated time required:	50 hours

The assignment should be submitted electronically as a pdf, in the form of a brief individual report (2500 words limit applies) explaining the methods and giving a clear statement of the principal results. Students should highlight the physical basis for any assumptions made wherever possible.

The final completed assignment should be submitted to Turnitin following the standard procedure on or before the submission date.

You are reminded that in the absence of exceptional circumstances (supported by written evidence) late submissions will be penalised.

Aim

The aim of this study is to select, apply and study network data analysis across different graph scales using passenger flight data.

Problem Definition

Many large data sets have a network dimensionality (social networks, peer-to-peer file transfer, bit coin, organisation, supply chain, connected systems…etc.). Practical applications of network analysis include understanding its resilience to disturbances, robustness to attacks, and efficiency of processes on networks (information, energy transfer). The techniques draw knowledge from applied statistics, applied physics, graph theory, and data science. More recent advances have developed graph neural networks to discover graph features to quantify challenging behaviour.

In this case you're to apply classic feature-based graph analysis to flight data to demonstrate multi-scale graph analysis covered in the lectures. The aim is to represent important notions in a low data dimension to enable effective communication and dissemination.

The objectives of this works are as follows:

1. Transform one month of flight data into a network form and visualise it using any software or coding package. You may have to do some minor data cleaning to ensure airport locations match and are reasonable.
2. Implement graph analysis across 3 different graph scales: 1) macro-scale (statistical analysis), 2) meso-scale (community analysis), 3) node-level (centrality analysis).
3. Assess and discuss the 3 different graph analysis scales and what they mean for air transport networks in real socioeconomic and engineering terms.

Data

The dataset is given on Canvas - contains domestic flight data for **1 month** across different countries in the world and different years. An example of the image from the database is shown below. Some degree of data cleaning maybe required to mimic real world data problems, e.g., the location of some airports maybe missing or incorrect.

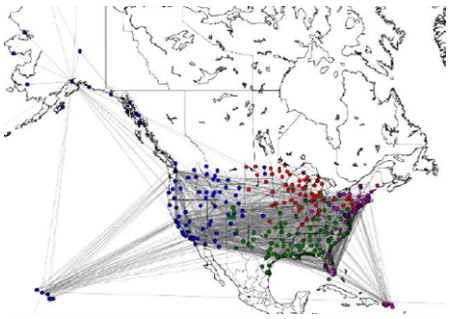

Figure 1 Example of domestic flight network for USA.

Methodology

First isolate a single month of choice in any given year and isolate the **4 countries of interest (internal flights only): USA, China, UK, and Australia.**

To achieve the objectives of this work you need to select various graph analysis methods across the scales (node to community to overall structure), justifying pros and cons in comparison with other methods (500 words max).

It is anticipated that one of the methods from the scope of the materials taught in the module will be implemented for analysis purposes. Any analysis tools are permitted – there is no requirement to use MATLAB/R/Gephi/Python or any other tool, but it is recommended to include the code in the report Appendix to enable markers to see possible reasons for the problems with the results (if this happens) to be accounted for during the marking process.

The word count for the main section of the report (not including Appendix and References) is 2000 words, therefore, results and discussion should be predominantly presented in the graphical or tabulated way. Reproduction of the problem statement and data description will not be needed.

Assignment report should be prepared using the supplied template.

Assessment

Assignment marking will be focused on the ability discuss the appropriateness of the graph techniques with the following selection and quality of the results assessment. A part of the marks will be allocated for implementation efforts and results obtained.

The marks for the assignment will be distributed as follows:

Q1. Construct separate airline networks for each country (internal flights only) and visualise them using geospatial coordinates provided. Comment on your how it is done. (250 words max)

[15 Marks]

Q2. Plot the following:

a) Degree distribution (x-axis: descending rank, y-axis: logscale of weighted degree). [5 marks]
b) Degree vs. betweenness distribution [10 marks]

c) Assortativity (degree-degree correlation) and state the value of Assortativity [5 marks]

d) Find a way to plot your method to calculate the Core community size (not modularity) [20 marks]

[Total: 40 Marks]

Q3. Please now reflect back on the analysis, discussion, and conclusions in Q1-Q2 previously. Consider why you selected certain metrics for analysis, and why others were not used. Consider the real-world aviation application of metrics for analysis and how airline operators might use them [10 marks]. Discuss how distribution results found in Q2 relate to population distribution, and airline/airport operations. [5 marks]

Hypothesize with evidence on the potential impact of results on the design of future aircrafts in these different countries by using random spatial graphs with configurable parameters. Analyse how will be changing the fuel price impact the distance penalty and the formation of random graphs that mimic the airline network? [15 marks]

[Total: 30 Marks]

Q4. Style and presentation that include presence of logical structure, appropriate citation style (if references used), quality of graphical material (labels, legends, titles references/citations, and captions as appropriate), readability of the text material, clarity of results presentation in the text.

a) Structure [5 marks],
b) Clarity [10 marks].

[Total: 15 Marks]