Network Analysis of journeys between stations within the London Underground network

COMPGW01: Complex Networks and Web
Coursework Project (2015-2016)

Sergiu Tripon
Student Candidate Number: KJFP0
MSc Web Science and Big Data Analytics

Word count: 5,000

Contents

1. Abstract

This study analyses the structure of the London Underground network from a geographical, day-based and time-based standpoint. The London Underground network is a crucial transport infrastructure with an exceptional impact on the UK economy. The study finds that the London Underground network is a small-world scale-free network, due to its high clustering coefficient and low average shortest path length. Furthermore, it discovers that each cardinal direction of London has at least one “hub” station that serves as the means to travel to Central London. Moreover, through the day-based and time-based analysis, the most frequent routes are presented and trends are also identified.

2. Introduction

2.1. Scope

This project aims to analyse and identify significant and interesting patterns in the London Underground journey information dataset using network analysis. I believe that this study will produce compelling results providing insights on specific trends emerging during certain periods of the day and the week within the London Underground network.

2.2. Research

TFL suggests two potential uses of the dataset [1] which I decided to consider:

- Create a visualisation of travel flows between all origin and destination points across the transport network
- Estimate journey times based on the sample journeys, taking into account the time of day

By combining TFL's suggestions with own interests and ideas, I compiled a number of questions and insights that the project aims to answer or achieve. Firstly, in terms of applying the network science knowledge gained thus far in the course, the study presents significant network metrics computed and verifies whether the network fits the "Small World" and "Scale-free" network types. Secondly, it analyses how London is linked together by the Tube. Thirdly, the project compares 7 different networks, one network per day of the week, in order to investigate whether any compelling day-based insights arise. Finally, it compares 4 different networks, one network per period of the day, attempting to identify interesting time-based trends that may emerge. The fundamental research questions that the project seeks to answer are listed below:

1. What are the significant network metrics of the network?
2. Is the network a "Small World" and "Scale-free" network?
3. How is London linked together by the Tube?
4. Do any interesting trends arise when comparing week and weekend journeys?
5. Do any interesting patterns emerge when comparing rush hour and non-rush hour journeys?

2.3. Value

Between April 2014 and March 2015, 1.3 billion passenger journeys were completed on the network, which is an increase of 3.2% in contrast to the previous year [2]. Considering that the number of journeys completed continues to grow each year, analysing a journey-based network is extremely beneficial and insightful. Moreover, it will provide reasoning of its structure and connections between its stations.

This project aims to analyse the structure of a journey-based networks, especially regarding time of day and day of the week. Separating the data in terms of time of day and day of the week and analysing the resulting networks proves to be significantly important. Besides providing a clearer overview of how the network's traffic behaves during these periods, it also offers further results such as average journey time.

As the majority of people tend to complete the same journeys at the same time daily, interesting patterns and trends appear. This study seeks to identify these trends while also justifying their behaviour.

2.4. Data, Algorithms and Tools

Transport for London (TFL) provides a data set comprising of a 5% sample of all Oyster Card journeys performed in a week during November 2009 on bus, Tube, DLR, TfL Rail and London Overground. [3] This project only takes in consideration Tube journeys, ignoring other journey data. Algorithms are used to compute the network metrics. This study focuses on metrics such as the degree, in-degree, out-degree, weighted degree, network diameter, clustering coefficient and average path length. During this study, several tools are used: Microsoft Access, Microsoft Excel, Gephi and NodeXL, Adobe Photoshop.

2.5. Results and Achievements

The main results of the study aid to understand the structure and flow of the network. Furthermore, the resulting insights also explain why the network is connected in a certain way. Finally, they also help discover new trends that showcase previously unknown network behaviour.

Moreover, other achievements involve finding answers to the research questions proposed. Firstly, it aims determine whether the network is a small-world and scale-free network. Secondly, the study seeks to identify insights by analysing differences between journeys completed during different times of the day or different days of the week. Thirdly, the study also focuses on understanding how London is linked together by the Tube. Equally important, the metrics computed will play an instrumental part in finding answers to the research questions of the study.

2.6. Self-assessment

I believe that the outcomes of the study are both interesting and valuable. In particular, analysing Tube journeys is something that I was always interested in and always had questions about. Working on this project offered me the opportunity to find answers to those questions through network analysis. The results of this study range from analysing and comparing journeys completed during different times of day or different days of the week to determine whether the networks is a small-world and scale-free network.

Two aspects of the dataset that I find very interesting to analyse are the time of day and day of the week journey information. Comparing frequent journey routes and journey times during the rush hour and non-rush hour periods provides interesting results that answer questions in regards to the network's behaviour and can also be used to estimate average journey times. Similarly, comparing journeys completed during the week and during the weekend, insightful results provide indication of the means of the journeys. For example, on Fridays, there are a lot of journeys from Canary Wharf to Waterloo which may indicate that a large number of businessmen working in Canary Wharf are travelling home, outside of London, using a train taken from Waterloo.

Overall, understanding how the London Underground network links London together is extremely engaging, because it answers questions in regards to the passengers' travel behaviour that other people may ask themselves regularly and that would remain unanswered without network analysis.

3. Background

3.1. Network Science

Network science is a research field involving the analysis and study of complex networks. Complex networks include Telecommunication networks, Computer networks, Biological networks, Cognitive and Semantic networks and Social networks. The unique actors within a network are known as the nodes, and represented on a visualisation by circles. The nodes have connections between them, which represent a relationship, and are known as the edges. Edges are represented on a visualisation by arrows.

3.2. London Underground

The London Underground network, known as the Tube, is a crucial service used daily to travel around Central and Greater London. The network is formed of 270 stations connected by 11 lines. TFL uses a zonal fare system to calculate fares. There are 9 zones, with zone 1 being the most central. This study focuses on Oyster card journeys only. It ignores paper tickets and Contactless payment cards journey, which were only introduced in 2014 [4]. The Oyster card is a smart card which can be topped-up using cash or debit card. Travel cards can also be loaded on an Oyster card. At the beginning and end of every journey, the Oyster card must be touched on the Oyster card reader, otherwise the journey is considered incomplete.

4. Literature Survey

Prior to establishing the scope of this project and deciding on the research question(s) it will seek to answer, a literature survey was conducted. The project benefits substantially from a literature survey. Firstly, it helps to identify related works to this study, which determines whether a similar study has already been done. This is extremely helpful because it helps to avoid replicating someone else's work. Secondly, it provides insights into what other authors have done such as datasets or types of analysis used. At the end of the literature survey process, 3 different studies were found which directly relate to this study, as they all analyse underground/metro systems from around the world.

The first academic work chosen is called *"TOWARDS A MULTI-MODAL SPACE SYNTAX ANALYSIS. A case study of the London street and underground network"* by Stephen LAW et al. [5] This study aims to understand the structure of London through developing a bi-modal network of London. The bi-modal network is validated through a statistical analysis with empirical data of station usage in Underground stations. The study uses London Underground movement datasets, which includes entry and exit figures from the different peak times for every station in London on both weekday and weekend. The paper's qualitative findings focus on the comparison of two models: bi-modal and street-only.

It found interesting results such as the segregation of South London from the Underground network and the locational difference between town centres with and without an Underground station. Additionally, the study identified locations such as Canary Wharf, which are segregated from the street network, but well-connected in terms of the Underground network. I believe that this research paper provides significant and successful results.

Furthermore, it is different compared to this study because it uses station usage data, and not journey data.

Furthermore, "*Network Centrality of Metro Systems"* by Sybil Derrible [6] focuses on applying the notion of betweenness centrality to 28 worldwide metro systems in order to find information that can help planners in their task of designing the systems of tomorrow. The study begins with a circle representation of every metro system. I thought this was very interesting as the differences in structure between each metro system were showcased. The research paper found predictive results with Brussels (the smallest metro) having the lowest betweenness centrality average while London (the largest metro) has the highest average. Further, the study continues by comparing betweenness centrality stats between the metro systems.

The key finding of the study is what the author calls "Democratization". The author experimented by adding new nodes, in order to see how the network adapts to the change. It was found that adding nodes spreads the share of betweenness across all nodes without favouring only a limited number of nodes (i.e. no ''winner takes all'' paradigm), which can be associated to a process of democratization, unlike degree distribution in scale-free networks for instance. In my opinion, this study is concise and insightful. However, I believe that some of the findings were slightly predictive, but the key results of the study will prove to be beneficial to scientists, planners and engineers.

Moreover, the third academic paper selected focuses on spatial network analysis and was published as *"Spatial Network Analysis of Multimodal Transport Systems: Developing a Strategic Planning Tool to Assess the Congruence of Movement and Urban Structure"* by Jan

Scheurer et al. [7] The common ground between the study I conducted and this academic paper is that both apply geographic and spatial network analysis to a transport system. As with the second academic paper presented, this study focuses on centrality and connectivity in urban public transport networks. This shows that centrality is the core and an essential focus when analysing transportation networks.

This study begins by introducing a GIS-based tool, designed to assess the centrality. The location focus of the study is Perth, Australia while the scope of the study is to analyse Perth before and after the Perth-to-Mandurah Railway was built. I believe this is an extremely interesting concept, providing clear contrasts of how the city was connected before and after. It may also highlight key benefits of the new railway being built, while also presenting its potential disadvantages or failures. The key results of this study show the railway's impact on accessibility, a 32% efficiency increase.

To conclude, I believe that performing a literature survey is beneficial to any author as it provides an indication of the topic's state-of-the-art. I believe that my study is significantly different from the literature presented because it focuses on journey data rather than usage data. The studies focused more on the properties of the networks instead of analysing traffic between the stations. However, I have certainly taken inspiration and was influenced by the literature presented.

5. Methodology

5.1. Dataset

The project uses a single dataset provided by Transport for London. It is available for free on TFL's website, although registration is required. The dataset comprises of a 5% sample of all

Oyster Card journeys performed in a week during November 2009 on bus, Tube, DLR, TfL Rail and London Overground [8].

Initially, the data was too large to import into Microsoft Excel, and it was imported into a Microsoft Access database. A query was designed to extract only Tube journey information. Furthermore, the query ignored station names “Unstarted” and “Unfinished” so that invalid station names would be avoided. The query result was imported into Microsoft Excel.

The data did not contain any geographical information of the nodes (stations), therefore a geographical-based visualisation of the network was not possible. In order to achieve this, Doogal.co.uk [9] was used to manually gather the latitude and longitude for every station present in the data. In contrast, the dataset consists of substantial journey information such as day, time, zone, fare etc.

Table 1: Dataset statistics

Nodes	266
Edges	42074
Network Type	Directed, Weighted
Average Degree	158.173
Average Weighted Degree	2552.94
Network Diameter	2
Graph Density	0.597
Modularity	0.253
Connected Components	1
Average Clustering Coefficient	0.741
Average Path Length	1.404

5.2. Algorithms

Clustering Coefficient

The clustering coefficient algorithm is used to identify clustering in a network. It is a numerical value between 0 and 1, with 0 indicating no clustering and 1 indicating high clustering. Clustering within a network occurs when nodes group together through a high proportion of edges between them. The clustering coefficient is commonly identified by searching for triangles. In a highly clustered network, when two edges have a common node, there is probably a third edge and the three edges form a triangle. This study uses the algorithm to determine if the network is a small-world network. Alongside a high clustering coefficient, a small average path length indicates that a network is in fact a small-world network.

Shortest path length

The shortest path length of a network is measure used to verify the network's efficiency. In this context, it identifies the shortest path between two stations. In order to calculate the shortest path length, the number of shortest paths has to be calculated first. Subsequently, the average shortest path length of a network can be calculated as the average number of edges through shortest paths considering all pairs of nodes within the network.

Closeness Centrality

> "An important node is typically 'close' to, and can communicate quickly with, the other nodes in the network." [10]

In order to determine how far a node is from all the other nodes, a sum of all distances from all other nodes to that specific node is calculated. This notion can be called farness. The closeness of a node was defined as a multiplicative inverse of the farness, therefore the more central a node is, the lower its total distance from all other nodes will be. Closeness centrality identifies important nodes by analysing how close they are, and how quickly they can reach other nodes within the network. Therefore, using this algorithm enables the identification of important stations within the network in terms of their closeness centrality metric [11].

Betweenness Centrality

"An important node will lie on a high proportion of paths between other nodes in the network." [12]

Betweenness centrality is used to identify nodes that connect a significant number of other nodes together. If a node is an intermediate between a high number of nodes, that specific node is deemed important. For example, to travel from Node 1 to Node 3, you have to go via Node 2: Node 1 via Node 2 to Node 3. If Node 2 appears recurrently, it is considered an important node of the network. If a high number of passengers use a specific station as an intermediate in order to get to their arrival station from their departure station, that station is an important one. Therefore, using and calculating the betweenness centrality metric of the nodes is extremely valuable [13].

5.3. Tools

During this study, the following tools were used:

- Microsoft Access - used to import the initial dataset and perform queries that extracted only data fragments required for the study.
- Microsoft Excel - used to import the query result and to apply different filtering (time of the day, day of the week). This tool was also used as the platform that NodeXL requires to run.
- Gephi - used to import the data, model the network, compute metrics, create graphs and perform most of the network analysis. Gephi's Geo Layout plugin was used to create the geographical visualisation of the network.
- NodeXL - used to calculate some specific metrics.
- Adobe Photoshop - used to design the visualisations exported from Gephi.

6. Results

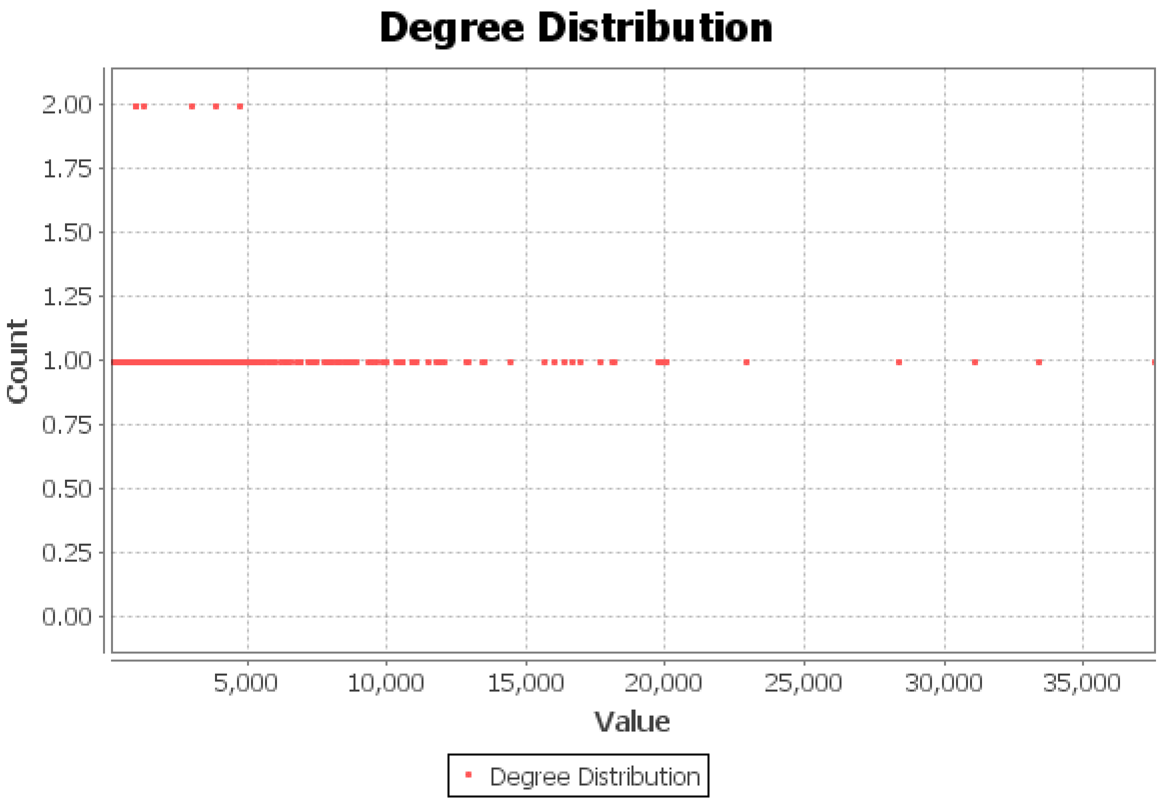

Figure 1: Degree distribution of the network.

6.1. Small-World and Scale-free

A network is a small-world network if the majority of nodes within the network are not close to one another, but can communicate with one another through a small number of edges. Social networks, the internet architecture, online encyclopaedias and gene networks are all known as small-world networks. In order to determine whether a network is a small-world network, the values of two independent metrics are used: clustering coefficient and average shortest path length. Networks that have a significantly high clustering coefficient and a low average shortest path length are deemed as small-world networks [14].

A network is a scale-free network if its degree distribution follows a power-law. This means that analysing any part of the degree distribution doesn't change its shape, and there will always be a number of nodes with a significant amount of connections followed by other nodes with a low number of connections. Many small-world networks are also scale-free [15].

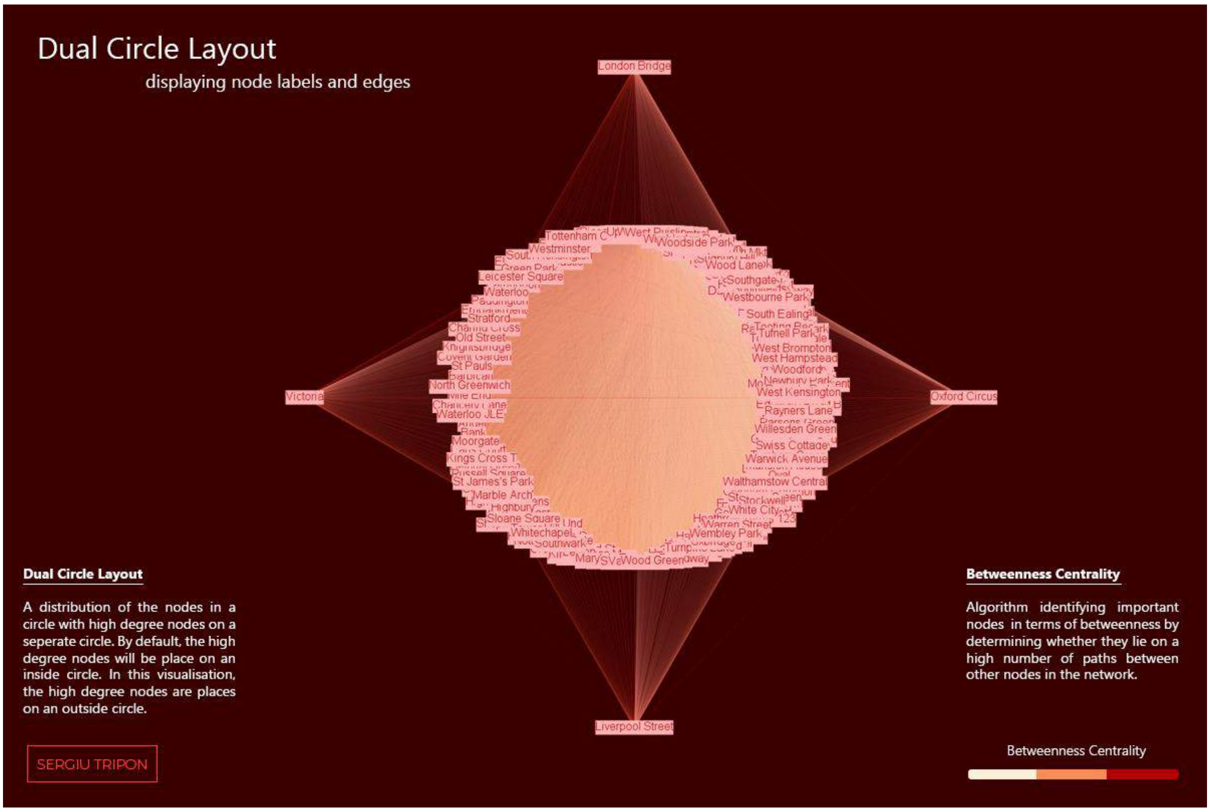

Figure 2: Dual Circle Layout of the network. Outer circle shows highest degree nodes: Oxford Circus, Victoria, London Bridge, and Liverpool Street. Full resolution figures are attached in the archive file.

By analysing the statistics from Table 1, the network studied has an average clustering coefficient of 0.741 and an average shortest path length of 1.404. As the clustering coefficient is a value between 0 and 1, the clustering coefficient of this network is considered significantly high. In contrast, the average shortest path length is low. The values of the clustering coefficient and average shortest path length indicate that the network is a small-world effect. In the London Underground network, the average shortest path length is the average minimum number of stations that passengers have to travel to get from one station to another. Unsurprisingly, in this case, the average shortest path is very low because the network is extremely well-connected in Central London and the majority of the journeys are completed within this area of the city. The clustering coefficient is considered as the probability of two stations directly connected to a third station, are also directly connected to each other.

Studying Figure 2, it is evident that only 5 nodes have a high number of connections while a high number of nodes have fewer connections. This indicates that the degree distribution of the network's nodes follows a power-law which further indicates that the network is scale-free. By analysing Figure 2, in the London Underground network, there are 5 different stations that are very rich in connections, with weighted degree figures ranging between 22,000 and 37,000.

6.2. The Tube linking London together

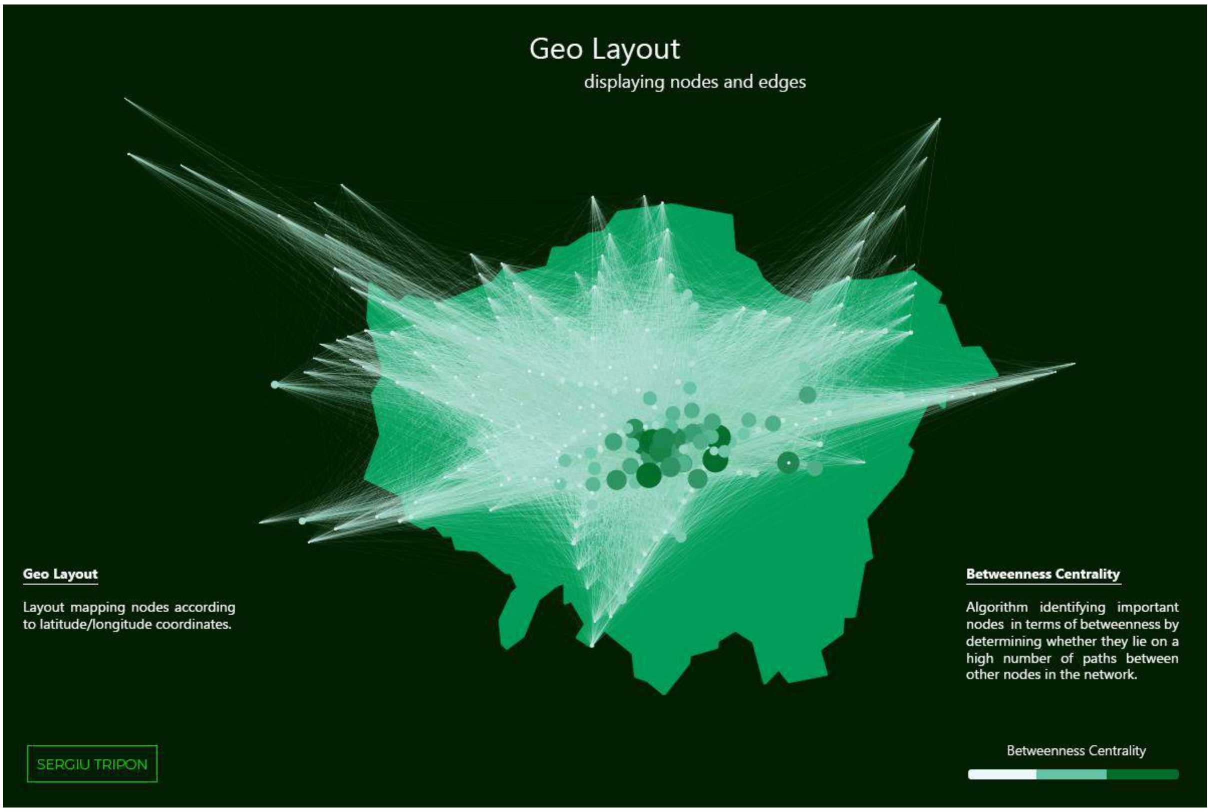

Figure 3: Geo Layout of the network, displaying nodes coloured and sized according to their Betweenness Centrality metric.

In Figure 3, Gephi's Geo Layout plugin was used to map the nodes which represent stations according to their latitude and longitude. This provides an overview of the stations' location and how passengers travel between them. The first evident observation is the lack of journeys in South East London. The main reason for this is DLR journeys not being considered for this study. The DLR is a major part of London's transport system and replaces large parts of the Tube network in South West London, compensating for the lack of Tube stations below the River Thames.

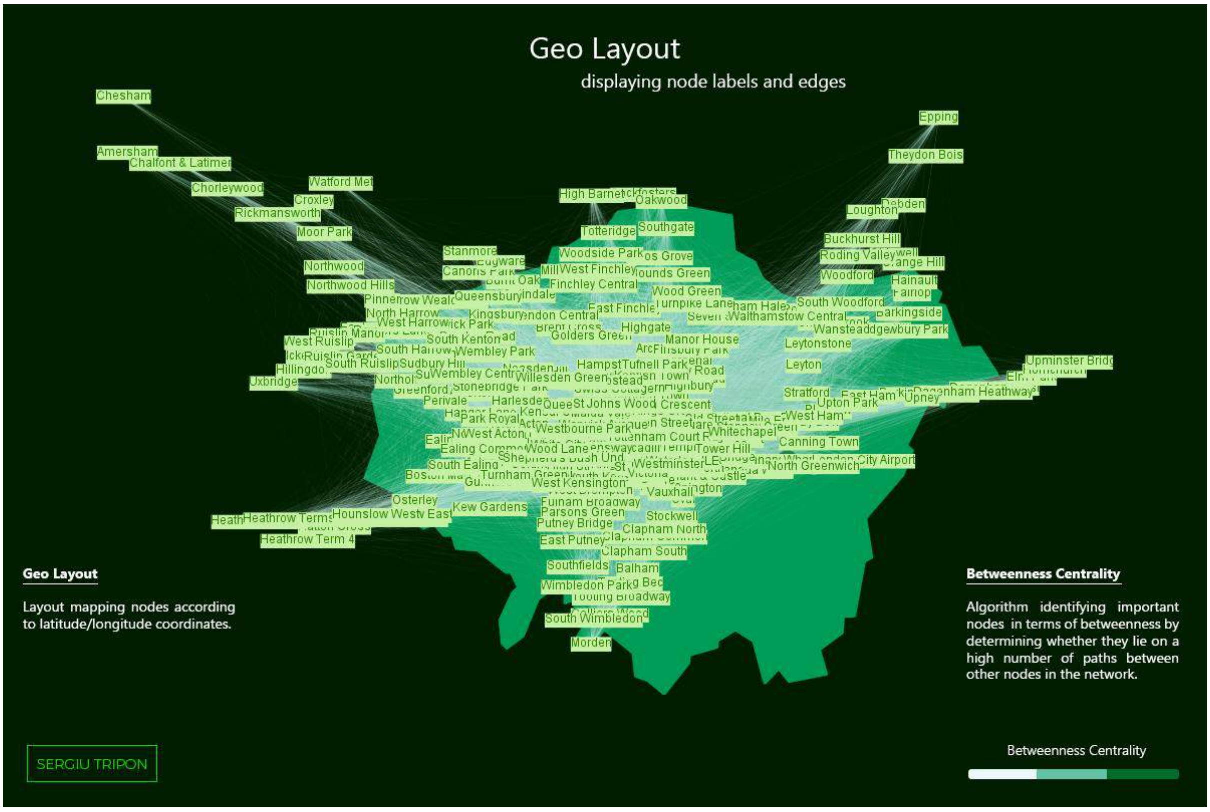

Figure 4: Geo Layout of the network, displaying node labels and edges.

The Betweenness Centrality algorithm was computed and used to shape the nodes by size. This shows the nodes of the network are used most frequently as intermediates between two stations. In the London Underground network, the nodes with high betweenness centrality represent stations in Central London. It means that passengers frequent stations in this area the most in order to get to their destination station. For example, in London, this is a realistic scenario because passengers travelling from North West London to North East London have to go through Central London. This means that stations in Central London will have high frequency, and therefore high betweenness centrality. Stations with high betweenness centrality are important within the Tube network. Furthermore, this means that Central London is used extremely frequently to get from one location to another. Subsequently, the evident question following is: How is Central London connected to other boroughs located in the outer parts of London, known as Greater London?

Analysing journeys involving parts of Greater London in Figure 4, it has been found that within each part of Greater London (North, East, South and West) there is at least one important node that is used as a hub to that particular area which connects passengers to Central London. In North London, Wood Green, Turnpike Lane and Seven Sisters stations link the North to Central London, while in East London, Stratford, Canary Wharf and North Greenwich link the East to Central London. On the other side, Brixton and Tooting Broadway have been identified as important hubs linking South London to Central London, while Harrow on the Hill, Wembley Park and Ealing Broadway are West London's important stations.

6.3. The difference between week and weekend journeys

Table 2: Most frequent journeys and stations with highest betweenness centrality by day of the week.

Day	Most frequent journey(s)	Stations with highest betweenness centrality
Monday	Waterloo - Canary Wharf London Bridge - Canary Wharf Victoria - Oxford Circus Oxford Circus - Victoria	Farringdon Canary Wharf Victoria
Tuesday	London Bridge - Bond St. Vauxhall - Oxford Circus Liverpool St. - Tottenham Court Road	Tottenham Court North Greenwich London Bridge
Wednesday	Liverpool St. - Oxford Circus Kings Cross - London Bridge London Bridge - Waterloo	Holloway Road Liverpool Street St James's Park
Thursday	Brixton - Oxford Circus South Kensington - Victoria Victoria - Kings Cross	London Bridge Oxford Circus Canary Wharf

Friday	Canary Wharf - Waterloo London Bridge - Kings Cross Oxford Circus - Brixton	Paddington Bond Street Victoria
Saturday	Canary Wharf - London Bridge Oxford Circus - Waterloo Tooting Broadway - Balham	Leicester Square Earls Court Waterloo
Sunday	Angel - London Bridge Victoria - South Kensington Leicester Square - Waterloo	Stratford Camden Leicester Square

Table 2 demonstrates the most frequent journeys and the stations with the highest betweenness centrality value for every day of the week. I believe that the most important emerging trend in the per-day analysis is the occurrence of the route Waterloo - Canary Wharf on Mondays and the inverse of the route on Fridays.

Canary Wharf is a well-known business district in East London, while Waterloo is Britain's busiest railway station by passenger use. Accommodation in Central London is limited and expensive, therefore a lot of people choose to commute from other towns and cities nearby. A realistic assumption would be that on Fridays, a large number of businessmen travel on the Tube to Waterloo station and then take the train to their homes, outside of London.

It can also be observed that some people choose to travel home the next day, Saturday, via London Bridge railway station. Hypothetically, on Monday, the businessmen return to London at Waterloo railway station, and then they travel by Tube to work in Canary Wharf. The next interesting step in the investigation would be to analyse if this trend translates to the time-based analysis.

6.4. Insights into the impact of the rush hour

The data was split into 4 parts: morning rush-hour, evening rush-hour, the afternoon period between the two rush-hour periods and the night-time period between the two rush-hour periods. The morning rush-hour is between 7.30am and 9.30 am, while the evening rush-hour occurs between 4.30pm and 6.30pm [16].

Table 3: Most frequent journeys, stations with highest betweenness centrality and average journey time by day of the week.

Time	Most frequent journey(s)	Stations with highest betweenness centrality	Average journey length (minutes)
Morning rush-hour	Brent Cross - Euston Tooting Bec - Temple Eastham - Elephant & Castle	Canary Wharf Latimer Road Borough	47.9
Evening rush-hour	Canary Wharf - Waterloo Waterloo - Canary Wharf Victoria - Oxford Circus	Piccadilly Circus Victoria Liverpool Street	36
In-between the morning and evening rush-hour	Upminster - Elm Park Maida Vale - Tottenham Court Road Blackhorse Road - Neasden	Regents Park Marylebone Shepherd's Bush Market	50.2
Before and After the morning and evening rush-hour	Clapham South - Bank Brixton - Canary Wharf Brixton - Chancery Lane	North Greenwich St. Paul's Barbican	45.1

In Table 3, the Canary Wharf - Waterloo and its inverse route appear in the time-based analysis just as they did in the day-based analysis. Interestingly, they both appear in the evening rush-hour period which causes some confusion. In the day-based analysis, an assumption was made in regards to this route. This assumption would be extremely realistic if the route Waterloo - Canary Wharf was frequent during the morning rush-hour as this indicates that the businessmen travelled back to work on Monday morning. However, the route appears in the evening rush-hour, which doesn't comply with the previous hypothesis as it indicates that the businessmen miss work on Monday. The trend's behaviour becomes slightly confusing and I believe there may be another reason for this behaviour.

7. Discussion

Through this study, it has been found that the London Underground system is a small-world network where the number of shortest paths through a specific station have distributions that are scale-free.

Furthermore, the study provided the visualisation and understanding of how Central London is connected. Second, it allowed the identification of important hubs outside of Central London and the analysis of how Greater London connects to Central London. It was found that each cardinal direction of London had at least one "hub" station that was crucial in connecting that part to Central London.

Moreover, a specific pattern emerged in the day-based analysis of the study showing the route Waterloo - Canary Wharf and its inverse as the most frequented on Mondays and Fridays, suggesting that passengers travelling those routes are businessmen commuting from their homes outside of London via Waterloo railway station. However, in the time-based analysis, it was found that both routes were most frequent during the evening rush hour, nullifying the assumption because if businessmen were travelling on those routes, they would be missing work on Mondays. This realisation added significant confusion to the assumption, which made it less realistic. In the end, it was concluded that there may be a different reasoning behind the behaviour of this pattern which could be the basis of future work.

Subsequently, the study also found that the lowest average journey time was during the evening rush-hour. This indicated that TFL potentially operates a more frequent and increased number of trains during the evening rush hour. More importantly, the second lowest average journey time wasn't recorded during the morning rush-hour but in the period before and after the two rush-hour periods. However, the difference is small, of only 2 minutes 8 seconds. In contrast, the average journey time difference between the evening rush-hour and the period before and after the two rush-hour periods was of 9 minutes 1 second.

I believe that this study produced a number of interesting and valuable results and insights. However, I also think that it had limitations. Firstly, the dataset was collected in 2009, 6 years ago. Evidently, it would be ideal to have a more up-to-date dataset which would enable a more current analysis and visualisation. Secondly, the dataset is a 5% sample. This limits the ability to gather a clear picture of the entire network. These limitations translate into potential future studies. I believe that the study would be increasingly different if a complete and more up-to-date dataset was available. Additionally, I think that an interesting idea would be to analyse specific London boroughs. This would allow us to study the area's connectivity and identify its advantages as well as what it lacks. I believe that this would be useful for local councils using the analysis to improve and adapt certain aspects of the borough's transport infrastructure.

Overall, I believe that this has been a successful project for two main reasons. Firstly, I believe that it has produced a number of significant results. And secondly, I am extremely interested in the project's topic and enjoyed computing different metrics, conducting the analysis and identifying interesting trends and insights within the network.

8. References

[1] Transport for London, "Legacy data feeds - Transport for London", *Tfl.gov.uk*, 2015. [Online]. Available: https://tfl.gov.uk/info-for/open-data-users/our-feeds?intcmp=3671. [Accessed: 30- Dec- 2015].

[2] Transport for London, "Facts and figures - Transport for London", Tfl.gov.uk, 2015. [Online]. Available: https://tfl.gov.uk/corporate/publications-and-reports/facts-and-figures?intcmp=28704. [Accessed: 30- Dec- 2015].

[3] Transport for London, "Legacy data feeds - Transport for London", *Tfl.gov.uk*, 2015. [Online]. Available: https://tfl.gov.uk/info-for/open-data-users/our-feeds?intcmp=3671. [Accessed: 30- Dec- 2015].

[4] Transport for London, "12 million contactless journeys made since launch - Transport for London", Tfl.gov.uk, 2016. [Online]. Available: https://tfl.gov.uk/info-for/media/press-releases/2014/november/12-million-contactless-journeys-made-since-launch. [Accessed: 07- Jan- 2016].

[5] Orca.cf.ac.uk, "Towards a multimodal space syntax analysis: A case study of the London street and underground network -ORCA", 2015. [Online]. Available: http://orca.cf.ac.uk/37616/. [Accessed: 31- Dec- 2015].

[6] S. Derrible, "Network Centrality of Metro Systems", PLoS ONE, vol. 7, no. 7, p. e40575, 2012. Available: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0040575. [Accessed: 31- Dec- 2015].

[7] Trove.nla.gov.au, "Spatial network analysis of multimodal transport systems: developing a strategic planning tool to assess the congruence of movement and urban structure: a case study of Perth before and after the Perth-to-Mandurah railway / Jan Scheurer, Carey Curtis. - Version details - Trove", 2015. [Online]. Available: http://trove.nla.gov.au/work/164317401?selectedversion=NBD49016903. [Accessed: 31- Dec- 2015].

[8] Transport for London, "Legacy data feeds - Transport for London", *Tfl.gov.uk*, 2015. [Online]. Available: https://tfl.gov.uk/info-for/open-data-users/our-feeds?intcmp=3671. [Accessed: 30- Dec- 2015].

[9] Doogal.co.uk, "London stations", 2015. [Online]. Available: http://www.doogal.co.uk/london_stations.php. [Accessed: 30- Dec- 2015].

[10] Sites.google.com, "An introduction to Centrality measures - An Introductory Course on Network Analysis", 2015. [Online]. Available: https://sites.google.com/site/networkanalysisacourse/schedule/an-introduction-to-centrality-measures. [Accessed: 31- Dec- 2015].

[11] Week 4: Slide set 10: Centrality.

[12] Sites.google.com, "An introduction to Centrality measures - An Introductory Course on Network Analysis", 2015. [Online]. Available: https://sites.google.com/site/networkanalysisacourse/schedule/an-introduction-to-centrality-measures. [Accessed: 31- Dec- 2015].

[13] Week 4: Slide set 10: Centrality.

[14] Week 2: Slide set 5: Small-world.

[15] Week 2: Slide set 6: Power-law networks.

[16] L. Porter, "Avoid Peak Times On The Tube", About.com Travel, 2016. [Online]. Available: http://golondon.about.com/od/londontransport/qt/commutertimes.htm. [Accessed: 07- Jan- 2016].