Mind the Gap:
Reuniting the Tube network with underlying spatial themes

Craig Foot[1], Pragya Agarwal[2]

[1]Deloitte & Touche LLP, 180 Strand, London
Email: cfoot@deloitte.co.uk
[2]Department of Civil, Environmental and Geomatic Engineering, University College London, London, Email: pagarwal@ge.ucl.ac.uk

KEYWORDS: graph, networks, nodes, spatial, usability

1. Introduction

In this project, graph theory was applied to the analysis of an urban transport network – specifically the London Underground. In modelling the network as a graph, a number of measures become available for use in an analysis of the network structure. The use of these measures will allow a comparison of the system with that of other world cities, as well as providing useful information for the planning of network expansion or the current running and maintenance of the network. A wide range of socio-economic indices also exist that are related to boundaries covering the spatial extent of the network. Using recent research by Agarwal et al. (2006), the project demonstrated an application of graph centrality indices in the context of London Borough boundaries in order to provide a link between a network analysis and data that is applicable to its spatial extent. It is expected that an improved understanding of the spatial structure underlying the tube network will enable enhanced usability of the transport network, support navigation choices in the urban area as well as aid future planning decisions.

2. Methodology

The first stage involved the development of a network model on which to perform the analysis. The aim was to produce a representation of the network in its simplest form – as an undirected graph with edges of equal weight. Social network methods were adapted to be applied in this particular case to a physical network and its underlying spatial structure. Measures of centrality given by graph theory were applied to this model, specifically centrality indices of degree, closeness, and betweenness. The degree of a particular node describes the number of nodes that are adjacent to it (connected by one link, or edge). The concept of betweenness centrality was introduced by Bavelas (1948) and formalised by Freeman (1977) in Equation 1.

$$c_B(i) = \sum_{k \neq j \neq i} \frac{S_{jk}(i)}{S_{jk}} \tag{1}$$

Here, the betweenness centrality of node i can be found as the ratio of shortest paths between nodes j and k that i is situated on to the total number of shortest paths to be found between j and k.

Another measure of centrality describes the how close a node is to other nodes in the network. In the context of transport networks it can be seen as an indicator of how accessible all destinations on the network are from a particular node origin. Sabidussi (1966) defines it as shown in Equation 2.

$$c_C(i) = \frac{1}{\sum_{j \in V} d(i,j)} \tag{2}$$

Here the closeness centrality of a node i is the reciprocal of the sum of all the distances between the origin i and every possible destination on the network represented by j, where j belongs to the set of nodes V. Distance in this sense refers to the graph theoretical distance, or the weight of each edge where weight represents the cost (e.g. distance, time or money) of traversing an edge. In an unweighted graph all edges have a weight of 1.

Data for the project was sourced from the London Tube map and from the online community in the form of contributors to websites such as OpenStreetMap. The positions of Tube stations are available from a number of sources, such as Ordnance Survey (OS) maps or Google Earth placemarks. A tabulated set of latitude and longitude values were located online and subsequently downloaded and imported into a spreadsheet. The locations were considered accurate enough for the purposes of the project, and a visual inspection of a sample of the data was used as a check. To ensure completeness, the downloaded coordinates were compared to the station names given in the station index of the January 2007 Tube Map. A number of versions of this map are available on Transport for London's website, in Portable Document Format (PDF), which allows for simple extraction of the complete list of station names from an official source. The station names were manipulated using a spreadsheet package, and matched to the latitude and longitude positions given in the downloaded dataset. A benefit of this method was that the zone attribute for each station is also included in the map index, and therefore was the first attribute to be added to the dataset. Other attributes can be easily added to the stations using the method of spreadsheet manipulation, and then visualised in a GIS such as MapInfo or ArcGIS.

Figure 1. Visualisation of the geographic patterns in the tube network

Network Analysis software traditionally employed for social network analysis was also investigated as a tool to represent and analyse the network. The values were normalized to the range 0 – 1 in order for future work to compare the Tube graph with other networks. Brandes & Erlebach (2005) was used as a reference for the normalization of the centrality indices. Figure 1 shows the geographic patterns of the London tube network. Connectivity in the tube stations, visualized in figure 1 as nodes, is analyzed as the degree of closeness in the stations. Figure 2 highlights discrete nodes on the graph that have a degree other than two, particularly highlighting the central concentration of high degree nodes and a peripheral group of lower degree nodes that provide a less dense zone of connectivity. This visualisation of the geographic dispersion of the transport network in London makes it apparent that the central parts of the city are better served and connected by the tube network than the outlying urban area.

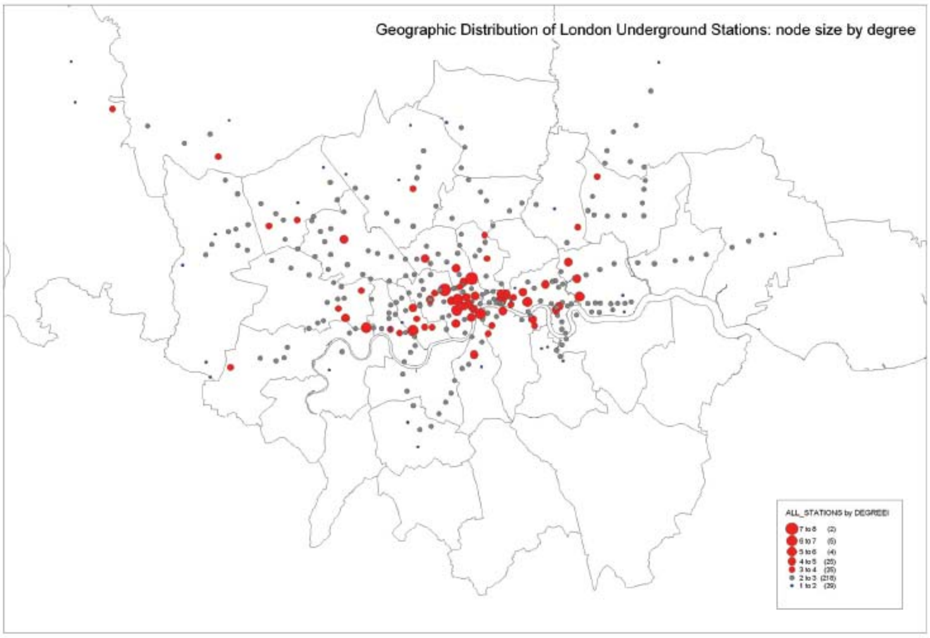

Figure 2. Spatial distribution and connectivity in the tube network

A different approach was also taken to relate the network data to the administrative boundaries. The concepts of geographic dispersion and heterogeneity after Agarwal et al. (2006) were applied to each borough where network data existed. A value for the geographic dispersion of each node was found and averaged by borough. Figure 4 shows an example of the thematic map output that this generated.

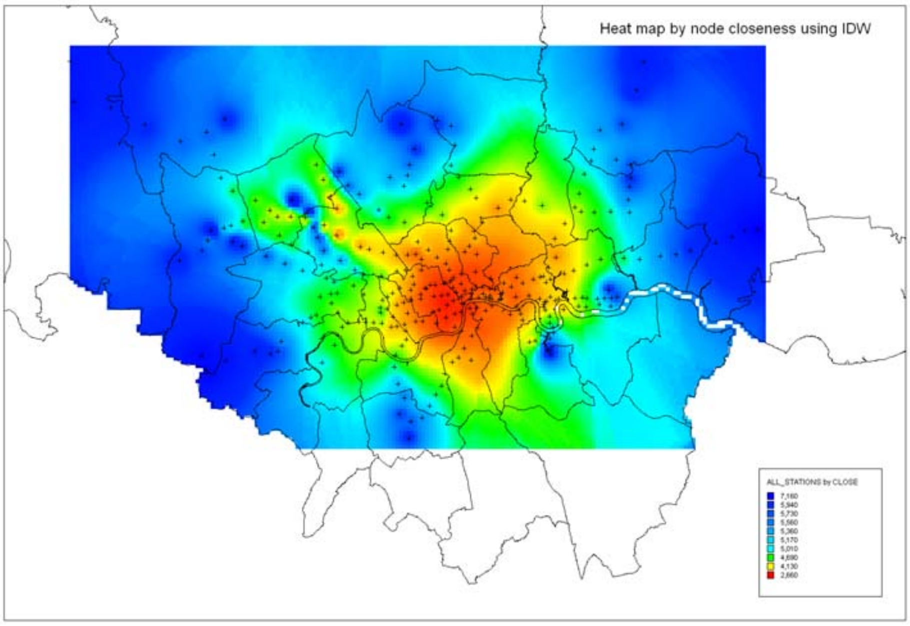

Figure 3. Closeness centrality of Tube stations

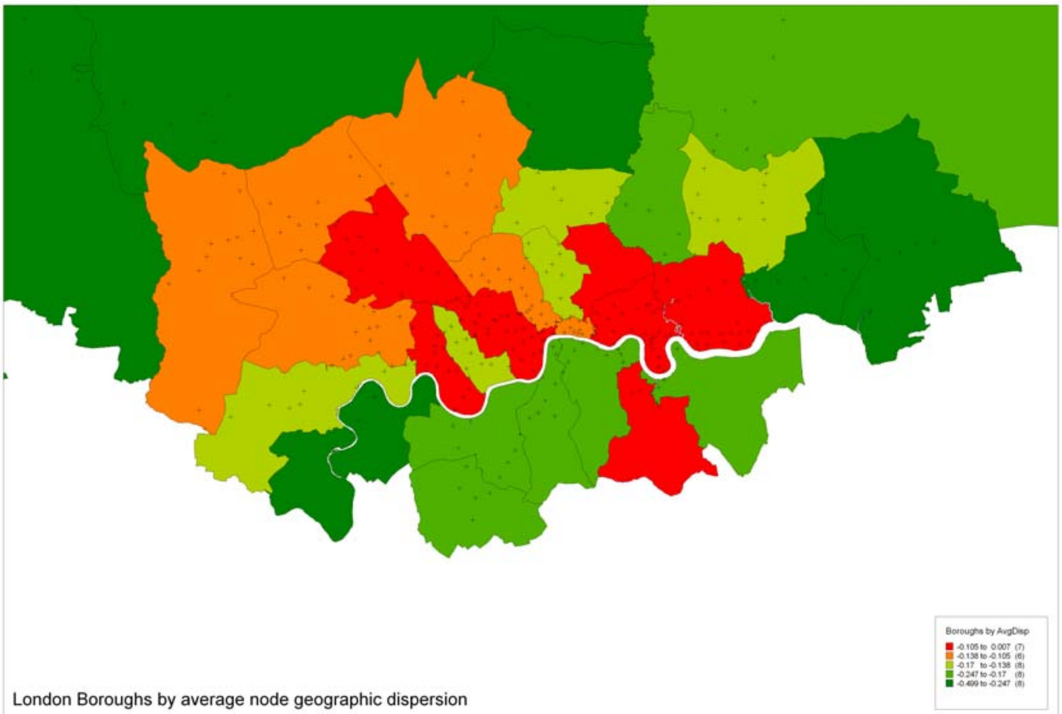

Figure 4. Average node geographic dispersion by borough

3. Analysis

The maps above provided some significant indicators. Firstly, the extent of the most central area can be appreciated as the warmest part of the image, and it is interesting to note that the strongest improvements to the closeness centrality of the system occur when Tube lines work in pairs, with one cutting out stations served by the other. This is illustrated by the fact that the mutual presence of the Metropolitan and Jubilee lines in the North West has produced local hotspots such as Harrow-on-the-Hill, Finchley Road, and Wembley Park. If the graph was weighted by travel time, the same effect would be achieved if lines were to run 'fast trains' that cut out intermediate stations – something that is only possible where the track permits it.

Cooler areas shown in Figure 3 indicate areas that provide the highest travel cost to other nodes on the graph. In the case of the Bakerloo line, poor centrality can be attributed to the high number of edges and few interchanges in this area. This highlights one of the first problems encountered with the over-simplification provided by the model: by not taking into account edge weights the closeness centrality equation is biased against sections of track that have a high station count. In fact, such sections may be closer to the rest of the graph in real terms if they are close together or a high speed is obtained en route. By weighting each edge by 1 it is assumed that the distances between stations, reflected in the travel time, is uniform throughout the network – which for the Tube is not the case. More data is required in order to weight the edges properly. The project's preliminary study was successful in collecting the track length distances between stations in Zone 1, after Jacobs (2002) and future work is going to look at extending the methodology developed here to analyse it.

The heat map for closeness centrality can be viewed as a starting point to highlight London boroughs that are good value for money in terms of travelling on the Tube. Such data could be developed to include a consideration of average house prices so that a query could return the set of boroughs or geographic areas that minimise house price or cost of living and maximise network accessibility. A heat map of the shortest path travel cost to a destination such as the workplace would highlight suitable origin locations that optimise travel cost in time, distance or money. An application of this information would be useful to a commuter who wished to live in an area that minimised the travel cost to work based on a set of constraints such as a house price or fare range.

4. Conclusion

The project has succeeded in providing a model of the London Underground as an undirected graph in its simplest form, providing a foundation dataset that allows future work to develop and reify as required for specific analyses. Data has been collated to provide a set of three centrality measures for all stations on the network: degree, closeness, and betweenness. This enables each station to be given a ranking based on each centrality measure and when joined with the list of geographic station locations, the distribution of centrality can be visualised.

The project has also provided a normalisation for each of the centrality measures so that the network can be compared firstly to networks of other transport modes, such as the London bus or tram network that shares the geographic space, and secondly to transport systems elsewhere in the world. The centralisation of the graph has been derived using normalised node centrality values so that an idea of network vulnerability – gauged in terms of its reliance on central nodes – can be clearly deduced. An additional benefit is that using normalised values to find graph centralisation enables easy, reliable comparison with other networks.

The concept of centrality has been extended to the geographic space by the creation of centrality heat maps. This decision in the development of the project was taken because extending network analysis techniques to a continuous geographic space, or one partitioned by boundaries, enables the application of GIS techniques to explore a wide range of queries that can be built and processed in a GIS, whilst taking advantage of proven data management techniques for spatial data. This approach combines well documented network analysis methods using graph theory with the flexible and extensible environment provided by GIS software to solve questions relating to areas such as location-allocation, access to culture, transport policy, and public security to name but a few. Each heat map consists of a grid of interpolated centrality values using the inverse distance-weighted method of interpolation. This process expands the discrete centrality values associated with the nodes to a more continuous space. The net benefit here is that the visualisation is improved when supplementary geographical contexts are overlaid. The centrality of the geographic space surrounding the graph has been further developed by the use of heterogeneity and dispersion indices to indicate the centrality of a set of partitions of the Greater London area. he example offered in this project was the use of Greater London borough boundaries, but others could have been deployed such as postcode districts, fare zones or isochrones radiating from a central point or destination.

5. References

Agarwal P, Béra R, Claramunt C (2006) A Social and Spatial Network approach to the investigation of research communities over the World Wide Web, *GIScience 2006, Lecture Notes in Computer Science*, 4197: 1-17.

Bavelas A (1948) A mathematical model for group structure, *Applied Anthropology (Human Organisations)*, 7: 16-30.

Borgatti S P (2002) *NetDraw: Graph Visualization Software.* Harvard: Analytic Technologies.

Brandes U, Erlebach T (2005) Network Analysis: Methodological Foundations. *Lecture Notes in Computer Science*, 3418: 1-471.

Freeman L C (1979) Centrality in social networks conceptual clarification, *Social Networks*, 1(3): 215-239.

Jacobs G (2002) *Railway Track Diagrams: 5 England South and London Underground.* TRACKmaps, Bradford-on-Avon. 64 pages.

Sabidussi G (1966) The centrality index of a graph, *Psychometrika*, 31: 581-603.

Biography

Craig Foot completed his BEng in Geospatial and Environmental Information Management at UCL in 2007. He is currently working for Petroleum Services at Deloitte where he assists with GIS development and related data management issues. His primary research interests lie in using relative adjacency and graph theory measures to analyse the effect of networks on demographics, the management of spatial data workflows and GIS usability.

Pragya Agarwal is a Lecturer in Geographic Information Science at University College London. Her primary research interests are in social and spatial networks, personalisation, location-aware systems and ontologies.