MEASURING THE STRUCTURE OF GLOBAL TRANSPORTATION NETWORKS

Sean Hanna
University College London
e-mail: s.hanna@ucl.ac.uk

Joan Serras
University College London
e-mail: j.serras@ucl.ac.uk

Tasos Varoudis
University College London
e-mail: t.varoudis@ucl.ac.uk

Abstract

The scale of analysis in Space Syntax has historically been limited to that of the city street network, with occasional studies of multi-city conurbations or comparative analyses of individual cities. We present a method for global analysis of the structure of international transportation, including roads, but also shipping, train and related networks. We assess the applicability of existing Space Syntax measures to graphs of this scale by undertaking a study of the international road and train network of Europe and its immediate surroundings, which we intend to be the first such study at this scale. While the closest such studies have focused on the effect of the network on movement at regional scales, finding less evidence for social and economic impact, we focus on international scale economic productivity as represented by gross domestic product (GDP), and observe that there is a strong relationship. This suggests that the structure of the transportation networks is an important indicator of economic activity at an international scale, and that such a methodological extension of Space Syntax is suited to revealing this.

Keywords: Transportation, Global, Networks

Theme: Modelling and Methodological Developments
Introduction

The scale of analysis in Space Syntax research has historically been limited to that of the city street network, partly due to the limits of computational resources, the availability of maps and data, and for theoretical reasons. This paper presents the initial stages of a global scale analysis of the structure of international transportation, including roads, but also sea shipping, train and related networks, which we intend to be the first such study at this scale.

Prior studies of street networks beyond the scale of an individual city have either involved the separate analysis of individual urban graphs, as in the comparative analysis of a set of cities (Peponis et al. 2007; Figueiredo and Amorim 2007; Hanna 2009), or have been limited to a particular region within a country, such as the regional mapping of connected cities in the north of England (Turner 2009). In part, this is a technical and methodological issue, of limited computational resources for larger scale analysis. It is also a theoretical issue, as many of the kinds of social phenomena—pedestrian movement, commercial presence, wealth and poverty, etc.—that can be observed in the relatively small scale of the city street do not have a clear presence at the much larger scale of an international road network, and more theoretical work would be needed to understand these (Turner 2009). Nevertheless, these studies have successfully revealed more abstract patterns. Turner (2009), in particular, demonstrated relationships between vehicular movement and measurements of choice in a road network at a regional scale, suggesting that at least some of the local findings of Space Syntax are replicated at a larger scale, and initiating a methodological framework for such study. This paper intends to further this line of enquiry, making contributions in methodology that may, in future, lead to extensions of theory.

As an extension of methodology, we propose the analysis of global transportation networks as graphs in the manner of street networks. In addition to roads, we include the international railway network, and leave for the future the possibility of including air and sea routes separately or in combination. The data for these latter are currently in preparation. Space Syntax measurements, of choice for instance, are highly dependent on the sight and movement constraints of the geometry properties of the network in space, which are expressed in the calculation as e.g. angular weighting. In graph analysis more generally, equivalent measures such as betweenness centrality are strictly topological. As scales increase, and particularly where movement between nodes is independent of the geometry of their intersection such as in air and sea travel, we believe that strictly topological measures will be appropriate, or that other cost factors such as fuel, taxes and route travel speeds may be more relevant. The quantification of these is also left for further research, but we anticipate that they may be accommodated by weighting of the graph.

To position this within a possible future extension of theory, we anticipate that such a methodology may allow the investigation of something analogous to the natural movement of pedestrians within an urban grid that is fundamental to Space Syntax, but at the larger scale of international trade and economic activity. Although the largest previous studies of road networks (Turner 2009) have concentrated on the network effect on literal movement because socioeconomic correlations appear to be nonexistent, these used as indicators the census data of wealth and poverty, which may have been at too fine a resolution. The work presented here uses much longer range networks and maps the relationship with GDP at an international level, and finds that such correlations become clear. While we are still some way from any theory of natural economic movement, our observations suggest that a portion of economic activity at an international scale is determined by the larger transportation network, and more specifically, by the network structure as revealed by the kinds of measures currently used in Space Syntax analyses at the smaller scale.
Method

This study uses road and train network data obtained worldwide from the Vector Map Level 0 (VMap0) which was released in 1997 by the National Imagery and Mapping Agency (NIMA) from the US. This release is the improved version of the map formerly known as the Digital Chart of the World, published on 1992. VMap0 is a 1:1,000,000 scale vector base map of the world. VMap0 has a transportation layer which includes the centre-lines representation of major roads and rail lines. The only classification given by the document specification of the data is that the road segments are categorised according to whether they are primary or secondary roads. Looking at the UK road map, it seems roads represented are mostly highways and some A and B roads. We are considering using more accurate global multimodal transport network representations datasets such as OpenStreetMap in future studies.

The analysis of the road and train network data has been performed in ‘depthmapX 0.24b’ (Varoudis, 2012). ‘DepthmapX’, originally developed as ‘UCL Depthmap’ by Alasdair Turner, is an open-source multi-platform spatial network analysis software that forked from the original. ‘DepthmapX’ can perform a number of spatial network analyses used in the research domain known as ‘space syntax’ and is used by a wide community of academic and practitioners. Space syntax is a set of theories and techniques, which apply graph network analysis to study the configuration of spatial networks in urban design and transport planning, and is originated from research by Bill Hillier and Julienne Hanson (Hillier and Hanson, 1984). In recent (Turner, 2007) space syntax network models, each street segment, or portion of straight street between intersections, is drawn to represent a node in the graph network. The road and rail centre line models extracted from VMap0 have been used to generate the graph structure for depthmapX.

Space syntax theorises the relationship between space and movement where a proportion of pedestrian movement is determined by the spatial configuration itself (Hillier et al., 1993). Empirically studies have found a strong correlation between the space syntax measure of ‘Choice’ with pedestrian flows and vehicular flows (Penn, 1998). Recent advances suggest that the use of angular segment choice explains aggregate movement better than topological or metric choice (Hillier and Iida, 2005). The choice measure in space syntax was first presented in (Hillier et al., 1987). Choice measures the quantity of movement that passes through each segment on shortest trips between all pairs of origin-destination segments in a system, and so is analogous to the mathematical graph measure of betweenness centrality that is widely used in network sciences. In detail, angular segment choice is calculated by summing the number of ‘angular weighted’ shortest paths that pass from a segment. More formally, choice (or betweenness centrality) measures how many times shortest paths overlap between all pairs of origins and destinations, where $g_{jk}(p_i)$ is the number of geodesics between node $p_j$ and $p_k$ which contain node $p_i$ and $g_{jk}$ is the number of all geodesics between $p_j$ and $p_k$ (Equation 1).

$$C_B(p_i) = \sum_j \sum_k g_{jk}(p_i)/g_{jk}(j < k)$$

Equation 1: Betweenness centrality from Hillier and Iida, 2005 referring to Freeman, 1977

Angular weighting of graph links implicitly associates a travel cost with turning direction; metric weighting represents the cost of distance travelled, and unweighted graphs capture network topology only. It is likely, as scales increase to the point where route choice decisions are assisted by maps, marked routes and GPS instead of visibility, that distance plays an increasingly more important role compared to turn angle, however factors such as terrain and its effect on travel speed are also relevant, and not presently available in the data. Due to similar lengths of segments in the map, angular weighting was used as a rough proxy for terrain and travel speed differences, and the space syntax measure of ‘angular segment choice’ was used throughout this study. This also allows for potential continuity with smaller scale studies of urban movement.
Initial tests of different measures on samples of the network did not reveal significant differences between weighting methods, however we feel this warrants more detailed study.

High values of choice indicate that a segment lies on the shortest paths for a high number of nodes in the system and it is significant to the function of the network as it can be said to be a ‘through segment’. During the analyses four cut-off radii were also set: 50km, 100km, 500km and 1000km. The cut-off is essentially excluding shortest path search for paths longer than the cut-off value in order to expose network structures that are more dominant at these scales. The number of segments (graph vertices) reachable from a source within a certain cut-off radii represents the ‘node count’ measure that source segment.

An issue to be considered while dealing with networks at a global scale is the role of map projections to minimise the error in the representation of networks. VMap0 uses the World Geodetic System from 1984 (WGS 84) which uses spheroidal reference surface to represent the network, however, depthmapX requires coordinates to be on a plane. For our analyses, we therefore projected the sample area into one. This is a relevant issue as it introduces changes in the angles from the original projection to the planar projection; this is visually apparent looking at the shape of England on the network figures from this article. Future studies will examine more accurate methods for translation of angles to better deal with 3-dimensional representations for very large sample areas.

We are also extracting Gross Domestic Product (GDP) in USD and GDP per capita in USD from the World Bank’s World Development Indicators data. It is also important to note that we have decided to use the values relating to the year 1992 so that the comparison between the network and the economic indices are as fair as possible. Even though the network data was released on 1997, we believe that the year for which one can account for good representation is still 1992. In any case, we also believe that the slow speed of change of the transport network itself will not lead to very different results using GDP values in a 5 year range.

**Analysis of the European Map**

The full study is working toward a complete global map, with terrestrial transportation within continents linked together by sea and air routes. A number of issues are still to be resolved with respect to the treatment of both sea and air transport, which currently exist only as relatively long range origin-destination pairs and so are very different from segments on land routes, and for this reason we focus in the current analysis entirely on the combined road and rail networks, restricting ourselves to a distinct geographical region in which these routes predominate. The portion of the global map containing Europe and the immediate surroundings (including a portion of North Africa) was selected for analysis. This has the advantage of continuity with other studies, and the availability of related data, used in the next section. The sample area for the transport network selected spans through 50 countries, however, in this analysis, we have only considered those countries whose network coverage was complete; in other words, some countries located at the boundaries of the sample area, were only partially represented by our map, and so were removed. We have removed 9 countries from the analysis leaving a total of 41 countries.

Where the two maps are used together, rail and road segments are treated as a single network with any two segments linked at their endpoints regardless of their type. This assumes the free movement from rail to road (or the reverse) at any such link, which, in the absence of more detailed information, approximates the most likely conditions at stations on the rail routes.
DepthmapX analyses of angular choice over the European region are displayed in Figure 1 and Figure 2. At first glance the network appears to behave similar to that of an urban region, albeit at a much larger scale. For relatively small radii of 50km, equivalent to the local neighbourhood of an urban centre, the map (Figure 1) resolves itself into clearly defined ‘hot spots’ representing major cities. London, Paris and Frankfurt are clearly marked, as are large conurbations such as the English West Midlands and around Cologne, Dusseldorf and Essen in Germany. At much larger scales of radius 1000km, equivalent to distances of international travel between countries, the network (Figure 2) is resolved into a clear foreground of major routes between these centres, and a generally homogeneous background. This is, at least in appearance, similar to the foreground and background networks seen in cities (Hillier et al., 2012).
This suggests that some principles of the grid may be universally consistent across scales. Where local urban through movement within a city network, as represented by choice at radii the scale of the neighbourhood (e.g. 1000m), highlight the centres of local retail and commercial activity, the equivalent ‘local’ scale on the global map (e.g. 50km) highlights urban centres that are likely to be the financial and industrial equivalents. Where through movement across urban networks, as represented by choice at radii approaching the full scale of the city, display major linear transportation routes, the same is true in the European map. The identification of major financial centres such as London and Frankfurt along these highlighted routes suggests that larger economic activity may correlate in a similar way, to be confirmed by economic data.

**Toward a global spatial economic map: economic analyses by country**

If a relationship between the transportation network and economic activity should prove to exist across scales, we would expect to be able to model and predict a variety of spatially located economic outputs, from city, to regional, to national scales. In the analyses that follow we test the assumption that measurements on the sum of segments within a particular region provide an indication of the economic output of that region. At present, our most reliable economic data is limited to the resolution of countries, and we test the plausibility of such a model with national GDP statistics.
Highlighted regions of high choice in the maps above would appear to correspond to zones of economic activity, indicating as a plausible initial hypothesis that the sum total of segment choice values within a given region correlates to its overall economic output as expressed by GDP. This was tested for all countries in the data set, correcting for the skewed distribution in the scale of countries by plotting each on a log-log scale. In doing so, two outliers are immediately obvious. Malta, with choice values far below any in the data set, appears to have a disproportionately high GDP. This is as should be expected given that, as a small, isolated island it has will not be appropriately connected by our transportation networks until sea, and possibly air routes are represented. Czechoslovakia, with a disproportionately low GDP, was in the process of dissolution during the year of the data (1992). Pending further analysis, these might be considered exceptions that lend credence to the rule. For the remaining countries, it can be seen that there is indeed a clear relationship between segment choice and GDP, with correlations of $r^2$ of 0.609, 0.584, 0.425 and 0.353 at 50km, 100km, 500km and 1000km scales, respectively (see Figure 3).

The distribution of choice values at these scales, however, results in very few segments of high value, with a vastly larger number of low choice ‘background’ segments, the latter of which actually dominate the overall sum. It is the number of total segments, rather than their values, that characterise the measurement of national total choice. This is evidenced by taking the correlations as above between a simple count of segments and GDP, in which a similar result holds with $r^2$ of 0.638, 0.611, 0.486 and 0.416 at 50km, 100km, 500km and 1000km scales. Two problems therefore exist with this basic measure: (a) it is dominated largely by the simple sizes of countries, and skewed by relatively few large ones, and (b) it indicates only a simple relationship between the size of a country’s transportation infrastructure and its economic wealth, while saying nothing about how the network is connected.
We address the first issue would ideally be addressed by a finer aggregation of economic data, but in lieu of this we can divide both the GDP and total choice values by the geographical area of each country. Figure 4 indicates the correlation between countries’ GDP per km² and total choice per km², resulting in clear correlations of $r^2 = 0.517$, $0.485$, $0.314$ and $0.231$ at 50km, 100km, 500km and 1000km scales. This normalisation has the secondary benefit of mediating some of the skew in data due to the uneven distribution of nation sizes, such that if plotted on a non-logarithmic scale, the same linear correlations rise to $r^2$ values of up to 0.719 (for 50km radius).

To say something meaningful about the shape of the network—where infrastructure is built and how it is connected—we examine the portions of the network that have values of choice over a given threshold. We exclude the bulk of the ‘background’ network that contributed to the general node densities above by counting only the top 10% of segments by choice value, and excluding the 90% below this threshold value. This results in a much more selective map that represents only the apparent urban ‘hot spots’ of the 50km radius map, or the prominent international routes of the 500km map. The count of these segments within a region therefore indicates not simple network density, but how well that region is connected within the network.

Plots of total GDP against total choice of the filtered networks are given in Figure 5. In this case the anticipated uneven distribution of the filtered portion of the network does not appear to justify the normalisation by unit of area (as above, Figure 4), but to mediate the uneven distribution of data we have again plotted correlations using log-log scales. These indicate that countries with more segments within this upper 10% choice band do indeed have greater GDP. Correlations are only slightly lower than those between total GDP and overall node density, with $r^2 = 0.587$ and 0.581 for choice radius of 100 and 500km. If the plots (Figure 5) are examined in succession, this correlation is strongest at 100 km and 500 km distances, which suggests that it is the major routes at distances roughly between international centres that may play the more significant role towards each countries’ economic ranking.

Continuing under the assumption that economic performance is strongly tied to network choice, the distance above and below the line of regression also indicates countries that might be considered to be over- or underperforming in their economic activity. The UK, France and especially Germany appear to be over-performers in this sense, while Poland and the Ukraine appear to underperform economically.
Both overall node density and locations of highest choice provide very good indications of a country’s economic standing in terms of GDP, but the two have different geographical distributions, and so provide two different stories. No specific causal direction can be definitively inferred in either case. Certainly it is plausible that overall economic capacity (GDP) allows more roads or rail lines to be built within a country, thereby increasing node density. It is also plausible that such density is a generator of economic activity. In the case of the relationship between high choice and economic activity, the suggestion that it is the much larger scale radii corresponding to international routes between countries, lends at least some weight to the likelihood that the network plays a causal role, as these routes span several countries and therefore cannot be created as a result of economic productivity of any single country. The attraction of neighbouring regions to a potential market, however, must certainly be acknowledged. In the absence of further evidence, it seems plausible that the transportation network and the international economy are both contributors to a cycle of feedback, with road and train links both creating economic opportunity, and being shaped by the outcome of this activity.
An east-west divide and natural economic behaviour

Figure 6: Map layout showing how we have defined countries belonging to Eastern (purple) and Western (green) Europe

Figure 7: Comparison between the average choice measures for road and rail networks against GDP per capita for Eastern (red) and Western (blue) Europe countries and for different radii

Many factors contribute to a nation’s total GDP, including agriculture and resources that might be highly dependent on land, so it is reasonable to expect larger GDP values for larger countries, and this is likely to be part of the explanation for the correlations between overall choice and total GDP as expressed (Figure 3). A nation’s population is also a contributing factor to overall wealth. Of possibly greater relevance than the total economic outputs examined in the previous networks.
section is the effect of the network on a country’s potential for economic activity relative to this population—how much opportunity does the network afford to each person, with respect to their position within it?

To answer this, we look at the GDP per capita of a country against the mean value of choice within that country. This follows methodologically from Space Syntax studies of urban movement assessing the relative count of pedestrians in each spatial unit of the city (e.g. Hillier 1993), in that we look at the distribution of economic wealth of each unit of the international population. The underlying hypothesis is that the greater choice levels within a given region indicates greater economic potential, and will therefore correlate with greater average wealth of each person living in that region.

Using the same set of countries as a sample, we initially find correlations with values of $r^2$ ranging between 0.31 and 0.36 for choice radii from 50 km to 1000 km. This is not completely insignificant, but much lower overall than the simple tests of total GDP above. There are, of course, a number of factors other than transportation networks that impact economic output, not the least of which is the difference in economic history of each country, so this should not be entirely unexpected. The road and train networks are persistent over a relatively long period of time, change slowly, and resulting choice levels may remain more stable than the fluctuations in any local or national economy. A potentially greater such factor, in the 20th century at least, is the east/west division of the former communist countries of Eastern Europe and free market economies of Western Europe, resulting in two very different and largely incompatible economic systems.

To control for this, we split our whole sample into two distinct sets for analysis (Figure 6 and Figure 7). It is clear immediately that these form two distinct clusters, linearly separable on a plot of GDP per capita against mean choice. It is perhaps not surprising that the cluster representing Western Europe uniformly outperforms that of Eastern Europe. More relevant, however, is the difference in how the network measures these two zones correlate with wealth. In Eastern Europe, no significant correlation can be found with economic output at any radius. Western Europe, by contrast, displays high correlations of $r^2 = 0.60$ and 0.61, peaking at radius 500 km and 1000 km. These are in the same range as those for overall GDP and node density, but indicate the potential economic productivity of each person in the nation.

It appears relevant that the radius distances at which the correlations peak are similar to the distances seen to be most relevant when looking at total choice values (above, Figure 3), and that these are roughly the distance between international centres. This implies that the range of impact of the configuration of the network is not limited to one particular country, but is an international effect involving organisation across regions.

To the extent this relationship between network and GDP is causal, it suggests that market economies in Western Europe have taken advantage of these transportation networks to engage in economic activity. Countries in Eastern Europe, by contrast, have not. It seems equally possible that in more controlled economies, there is not such opportunity for the network to guide economic activity by allowing markets to freely self-organise, resulting in the absence of correlation with GDP. Where natural movement of pedestrians is the movement determined by the configuration of the street grid in absence of other limiting factors (Hillier 1993), we would propose here that there is a form of larger scale economic behaviour determined by the much larger scale network we have examined. Although the details and mechanism cannot yet be discerned, the observation suggests that there is at least an economic activity that is ‘natural’ inasmuch as it is free to exploit the network capacity, and therefore determined by this only when external political control is absent.
The details of international scale economic activity may be quite different from that pedestrian movement. Movement is unlikely to be based on sight and the cost of angular turns, but on the topology or different weightings of travel within the network. Transactions will be based on longer term planning than immediate colocation. Also, phenomena have yet been observed only against economic, not other social factors. Each of these can be tested empirically. Regardless of these differences, it would appear that the topology of transportation networks has a clear capacity to order behaviour, to the extent that agents are otherwise free and self-organising, at multiple scales, and further research is warranted at the level of global economy, as has already been valuable in these other systems.

Conclusions

The analyses here are presented as an initial study of the feasibility of such a global scale analysis, but indicate that some of the phenomena noted in smaller Space Syntax work do apply at this level. GDP has been used as the primary economic indicator, aggregated at the scale of individual countries, but considered both as a total and per capita. GDP may be considered an indicator of the general state of national industry. This was found to correlate well with the total number of segments within a country, and also to a threshold number of high choice segments at radii between 50 km and 500 km. While the former may appear to be caused by economic activity, the scales at which choice routes appear suggests at least some feedback at a regional level.

Per capita GDP may be considered the ability to maximise each individual person’s output, so a relationship to the network would appear more likely to be effect of the network on social economic activity. Strong correlations were observed with national mean choice, this time peaking around 500km and 1000km, but only for the countries in Western Europe, while eastern countries displayed no correlation. We suggest that this is a result of the influence of the network on free economic behaviour, which in the western countries has historically been allowed more freedom to markets constrained only by the physical network. We anticipate that a similar study of historical economic data will help to refine this, and plan this in the near future.

The context of the work in this paper is in part a much larger study of global dynamics, including trade, migration, security and development aid. The ENFOLDing project is an interdisciplinary collaboration that seeks to model these and their effect on one another, as the complexity of such interactions is little understood but crucial to international relations. In providing a basis for understanding at least the economic effect of physical infrastructure, it is our hope to allow further understanding of global social phenomena such as trade and development. The measures used here have counterparts in the graph methods used in economic analysis (Bloehl et al. 2011) such as the input-output models of different sectors of an economy, also represented as a graph. In these, measures of centrality have been used to understand the structure and behaviour of economic movement completely decoupled from space, but it appears to be methodologically possible to begin to analyse a combined network of spatial and economic transactions at several different levels, and we are currently working to do so. The next phase of analyses will include networks we are assembling for air and sea transportation, which should allow the separate industries represented by these economic models to be unpacked when compared against each relevant mode of transportation, and better understood in a spatial context.

We anticipate that the coupling of the spatial system with political and economic networks will show promise for the extension of these methods. An understanding of this network, and its most likely routes of concentrated activity as revealed by measurement of high choice or similar,
would indicate points most susceptible to shock, or intervention. Scenarios for disruption of the network can be tested, against known disasters, or development. As such we would hope to provide understanding of phenomena crucial to international relations and policy decisions.

References


