Space Syntax and Spatial Interaction: Comparisons, Integrations, Applications

Michael Batty

CASA, University College London,
90 Tottenham Court Road,
London W1T 4TJ, UK

m.batty@ucl.ac.uk

1 February 2017

ABSTRACT

In this paper we attempt, somewhat optimistically perhaps, to compare space syntax with spatial interaction. At one level, these two approaches to urban spatial structure are non-comparable. Space syntax is largely a descriptive technique for visualising spatial relations at the level of connections between places while spatial interaction is a predictive model that forecasts how much travel there will be between places. Space syntax articulates the system in terms of whether or not a physical link, usually at the level of the street system, exists while spatial interaction predicts movement between all origins and destinations which are places usually anchored in terms of the street network, but which at the level of prediction assume connections between all places. Space syntax tends to be grounded at a fine spatial scale while spatial interaction defines places as aggregates of activity in much larger zones than the scale of the street system. The main output of space syntax is a connectivity matrix of step lengths between streets whereas in spatial interaction, such networks are predetermined, measurable in terms of Euclidean distance or generalised cost of travel, while the output is the volume of travel prior to this being assigned to the relevant modal network. However both approaches define accessibility, in space syntax in terms of streets while in spatial interaction in terms of places or zones.

There is however a fundamental way of relating the implicit network graph of spatial interaction to the explicit planar type graph of the street network by assuming the planar graph of the network is conceived of as a primal problem of spatial interaction while the dual graph linking streets in the planar graph is the graph which is used in space syntax. We exploit this duality and show how we can move easily between spatial interaction and space syntax from primal to dual and back again. This is grounded in a more fundamental graph which is the bipartite graph or list of streets/arcs and their intersections/nodes from which the primal and dual emerge naturally. We explore various distance and accessibility measures and show how they relate and correlate. We then go one step further and consider how various processes of random walking take place in these networks and look at the steady states of the primal and dual problems in terms of the likelihood of a random walker visiting any node or street. We define primal and dual Markov chains that enable us to generate these probabilities although there is some controversy as to whether higher values are associated with more important nodes. Nevertheless, this provides a basic framework for comparing primal and dual in comparing spatial interaction with space syntax. We illustrate these measures on simple and easy to articulate graphs and then extend this to a synthetic network of nearest neighbour links in Greater London based on some 699 nodes and 1972 symmetric street or routes/links between zones. We also speculate how we might begin to compare the predictions from related spatial interaction models with the street accessibility values from space syntax and in doing so, suggest that that there are ways forward in comparing outputs at the level of movement on links, notwithstanding that these two approaches exist at different level. This paper is a preliminary attack on the problem of linking these two approaches which remains problematic.

Preamble: The Problem Context

Space syntax is a descriptive technique for working out the relative accessibility or nearness of a set of spaces, often defined as streets, to one another with the purpose of comparing their relative nearness to the movement associated with each space or street. The assumption is that movement increases linearly with accessibility. Accessibility is often called integration. It is defined by first identifying a set of streets which are usually lines of unobstructed movement, sometimes called axial lines, and then observing whether or not any one street is connected to any other which is defined if any two streets are connected (Hillier, 1996). The set of links between streets forms an interaction matrix which can be viewed as a topological or binary graph on which various operations can take place, for example, to find shortest routes from any link to any other which are then used to define the relative importance or accessibility of one link to any other. When these accessibilities are summed for each particular link, this gives the relative accessibility of a link. One of the features of space syntax is that it works with binary (0-1) links between streets and Euclidean distance or cost is not a feature of the analysis. The model is only predictive when the accessibility values for each street are associated with observed movement in each street which is occasionally used to build a predictive model.

Spatial interaction models, on the other hand, predict movement directly between a set of locations that can sometimes be interpreted as intersections between streets but are usually more generic – centroids defining a location or area – and apply at different spatial scales. Interaction or movement is directly proportional to activities that are located at different locations and inversely proportional to some measure of the length of the street measured as some generalised metric incorporating physical distance, travel cost and/or travel time. The model thus predicts movement as a function of these independent variables. Accessibility measures consistent with the model's predictions can be derived but these pertain to locations, not streets or links between these locations. The models are parameterised in such a way that movement is estimated to be as close as possible to observed data flows.

There are several key differences between space syntax and spatial interaction. Space syntax is essentially a descriptive measure of street accessibility which is related to movement in a comparative rather than predictive way. It is not parameterised and as such, there is no estimation or calibration procedure used to operationalise the model. There is nothing in the technique that generates movement as in spatial interaction models. Space syntax does not deal with locations but with links between locations – streets – which in turn are defined as linear spaces. Links between these spaces are represented not in terms of distance either but as logical links which define a topological network. And space syntax does not incorporate any measures of activity associated with locations. The independent variables in spatial interaction models however are measures of (trip-making) activity at different locations, generalised distance between locations, and parameters that define the relative weight of these activities and distances. Space syntax is more parsimonious being based on logical links between spaces and forming accessibilities from these. Its only independent variable is the defined topology of the links which in some instances have been extended to other geometric properties such as street orientation (angular variations). These might be parameterised (weighted) but there are few if any examples which follow in this direction. After the model has been built, accessibilities are then compared with movement; if a strong linear relationship exists, then occasionally the model has been used to predict movement, usually in situations where new street links are added as part of a design.

The key link between the two types of model relates to the underlying network between locations. Both techniques begin with the physical network. Spatial interaction models predict flows directly from this and other locational data. Space syntax works out accessibilities from this network and then compares these to real data and if the correlation is good, a linear model can be fitted and used for prediction. The common key is the network but there the similarity ends. In spatial interaction, a network is defined as a set of intersections between segments – nodes and arcs – which is measured by some generalised distance or cost. The network does not privilege nodes or arcs in any particular way as this is the assumed backcloth on which spatial interaction takes place. In space syntax, one begins with the same network but from these, sets of segments that have their own integrity are defined as sequences of links. These form 'streets' and are in general composed of more than one segment. Once these streets have been defined, space syntax defines connections between streets as the existence (1) or not (0) of a node in the original network that defines whether or not the streets are connected. In some sense, this is the dual of the original network which we might think of as the primal but as streets can be constructed from more than one segment, this need not be a strict dual. We will however refer to the ***space syntax network as the dual*** and the ***original street segment network as the primal*** for in some senses, this is the key difference contained in the distinction between the two approaches. In fact, whether one uses the dual or the primal in space syntax does not make that much difference to the ultimate computations of accessibility (see Batty, 2013 for a detailed explanation and comparison) but it is not our main purpose here to focus on these empirical differences and similarities. Our quest is to see how close the two models are and how one might be linked to the other.

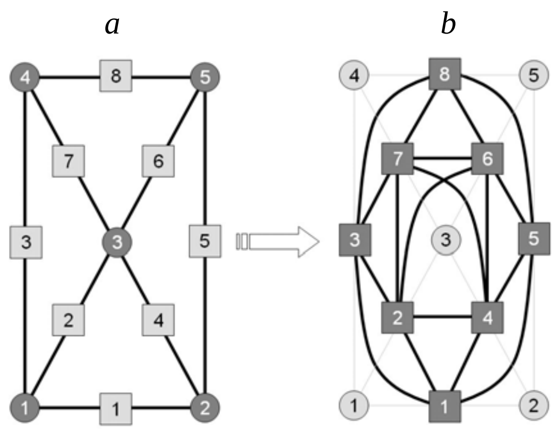
In the sequel, we will begin with the common key to both space syntax and spatial interaction which is the network. We will first develop a generic representation of the network from which the space syntax and spatial interaction variants can be defined. In essence, we invoke the idea of a pre-graph – a bipartite graph linking intersections/nodes/junctions or zone centroids to segments/links or arcs that are some form of route such as streets. From this bipartite graph, all else can be derived but it is important to note that these tools and models are a limited set of possible forms that can be defined as graphs, and the way its links can be measured. Other conceptions and variables based on locations associated with an underlying generic network can be defined which are not directly related to space syntax (Marshall, 2015).

The Generic Representation of the Network

Networks in space syntax and spatial interaction are usually embedded in two-dimensional space defined by locations which are points where street or route segments intersect. In general, these networks are planar graphs although this can be complicated if the representation extends into the third dimension, includes one way movements, or segments that cross one another without intersecting. In this context, we will deal exclusively with networks in two dimensional space where distance is measured using Euclidean geometries and the graph is planar (Barthélemy, 2011) The basic common ground between these two approaches is essentially a structural network of locations and paths of movement between: it is a graph whose elements are unweighted binary links. There is nothing else which is common to the two models and space syntax only uses this graph to derive any and every kind of prediction and insight from the model. Spatial interaction takes this graph as a skeleton network, loads or weights it with Euclidean distances or costs along its segments, then if incomplete, works out shortest routes to produce a full distance or travel cost matrix and then proceeds to use this as one of the basic inputs to the model that predicts flows between locations. In short, space syntax is very different from spatial interaction and it might be supposed that there is little point in trying to compare these as their basic networks of interaction are not the same. But as we will see, there is some interest in making a comparison because both models deal with location and movement, and thus it is worth attempting to see how spatial interaction ideas can inform space syntax and vice versa.

To provide the requisite intuition for the problem, let us propose a hypothetical planar graph of a street system where the nodes are the intersections between the lines (arcs or segments) which are the streets. In Figure 1 (a), we show such a street network which is composed of $N = 5$ intersections and $L = 8$ streets. This network is highly simplified: it is symmetric, that is the graph $G(N, L)$ is non-directed and there are no self-loops. We have not specified any weights for the links in this graph and thus it relates only to the system's topology.

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$



$$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Figure 1: The Primal (a) and Dual (b) Graphs and Adjacency Matrices of an Hypothetical Street Network

Clearly the planar graph which is Figure 1(a) links the 8 streets through 5 intersections which are numbered by the dark circles with the solid line representing street segments. In Figure 1(b), the streets are numbered by the dark squares and a solid line is drawn if two streets are connected through a common intersection. In the primal problem, each street has no more than two intersections where it joins other streets while in the dual each street can only intersect once with another street. It is this that can be relaxed in space syntax where a single street segment can intersect with several different street segments and this changes the nature of intersections. In fact, in this paper, arguably the graphs we define, which do not allow a street to have more than 2 intersections, miss some of the key elements of urban structure but it is a generic criticism of space syntax anyway in that the starting point is always a planar graph of local or nearest neighbour links. If the planar graph has N nodes, there are always many less than N^2 links, that is $L << N^2$.

What we require is a method for building primal and dual network graphs from the same basis and to this end, we begin with the planar graph and a list of nodes that are associated (or not) with a list of streets. The starting point for both techniques is thus the skeleton network of links between N nodes each of which we will define as $i = 1,2,3,...,N$, and L lines (segments or arcs) defined as $j = 1,2,3,...,L$. The basic representation is given by the matrix $\mathbf{A} = \{A_{ij}\}$ which is a binary matrix where $A_{ij} = 1$ if node i is linked to line j and $A_{ij} = 0$ if no such link exists. From this matrix, we can define the two basic matrices used in spatial interaction and space syntax. First the skeletal spatial interaction matrix which we call the ***primal*** counts the links between nodes as

$$U_{ik} = \sum_j A_{ij} A^T_{jk} \tag{1}$$

where the transpose operator T reorders the basic matrix $\mathbf{A}$ as $\mathbf{A}^T$. The space syntax matrix is the ***dual*** of this operation and is formed as

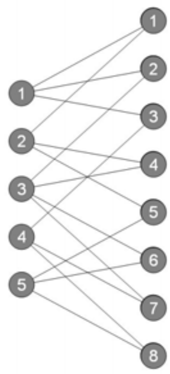
$$V_{j\ell} = \sum_j A^T_{jk} A_{k\ell} \quad . \tag{2}$$

Expressed in matrix notation equations (1) and (2) can be written as $\mathbf{U} = \mathbf{A}\mathbf{A}^\mathbf{T}$ and $\mathbf{V} = \mathbf{A}^\mathbf{T}\mathbf{A}$.

To provide some sense of what the primal and dual of these operations means, it is best that we introduce a simple worked example where we anticipate that the spatial interaction matrix is one where each line links two and only two nodes i.e. there are only ever two nodes which represent the beginning and end of each line, while the space syntax matrix (associated with this spatial interaction problem) is the opposite – the dual – where each line is now considered as a node and from two such nodes, there is only one line. To derive the general space syntax problem we need to relax this requirement but to produce the clearest example, we will adopt this simplification, and there is no loss of generality in proceeding this way. Now the matrix $\mathbf{A}$ can be graphically displayed as a bipartite graph (Borgatti and Everett, 1997) where we link nodes to lines – street intersections to streets. The example we used in Figure 1(a) has $N = 5$ nodes and $L = 8$ lines whose matrix and bipartite graph are defined as

$$A_{ij} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

and

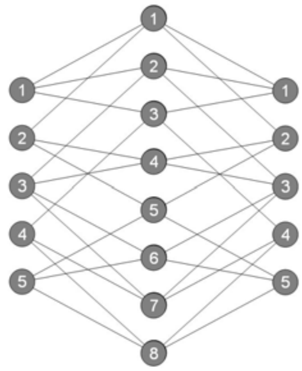


Note we define the 5 nodes as 1, 2, …, 5 and the 8 lines as 1,2, …, 8 with no ambiguity from the above definitions. The matrix, as we have been at pains to point out, simply records which street node or street intersection is associated with which each street line.

The two spatial models that we are examining take this information and deal with it in consistent but different ways. The spatial interaction model works by defining a matrix of interactions between intersections which are assumed to be centroids around an areal location and uses this matrix to predict the amount of movement. Then from equation (1) using the above bipartite graph we form the interaction matrix $\mathbf{U}$ as

$$U_{ik} = \sum_j A_{ij} A^T_{jk} = \begin{bmatrix} 3 & 1 & 1 & 1 & 0 \\ 1 & 3 & 1 & 0 & 1 \\ 1 & 1 & 4 & 1 & 1 \\ 1 & 0 & 1 & 3 & 1 \\ 0 & 1 & 1 & 1 & 3 \end{bmatrix}$$

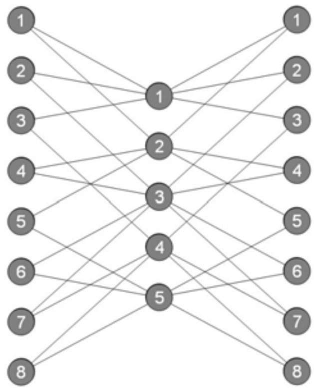
and



This is the primal problem. The dual involves defining how each street line is connected to any other, thus forming an interaction matrix $\mathbf{V}$ between street lines rather than street nodes and this defines a related graph. Then from equation (2) we get

$$V_{j\ell} = \sum_j A^T_{jk} A_{k\ell} = \begin{bmatrix} 2 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 2 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 2 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 2 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 2 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 2 \end{bmatrix}$$

and



If we compare the matrices $\mathbf{U}$ and $\mathbf{V}$ with the left and right matrices in Figure 1, then we can easily see that these are the same except the main diagonal elements of each of these matrices are equal to zero. In fact the main diagonal element reflects the number of paths in the graph to get from one node to the same. In the definition of $\mathbf{U}$ above, you can see that there are 3 steps to go from node 1 to itself and so on which are displayed from the juxtaposition of the two bipartite graphs and matrices $\mathbf{A}$ and $\mathbf{A}^T$. For $\mathbf{V}$, there are 2 steps to get from node 1 to node 1 via two lines and so on.

One final step remains to get the skeletal configuration matrices used for the two models and to do this, we need to slice out any links with more than one path and get rid of the self-links. Then the two matrices in question which define the primal and dual problems can be formed as follows:

$$\left.\begin{aligned} X_{ik} &= \begin{cases} 1 \, if \; U_{ik} \geq 1, i \neq k \\ 0 \; if \; U_{ik} > 0, i = k \end{cases} \\ Y_{j\ell} &= \begin{cases} 1 \, if \; V_{j\ell} \geq 1, \; j \neq \ell \\ 0 \; if \; V_{j\ell} > 0, \; j = \ell \end{cases} \end{aligned}\right\} \qquad (3)$$

It is quite clear that if these operations are accomplished, the matrices $\mathbf{U}$ and $\mathbf{V}$ are sliced to remove the path lengths and ensure that the matrices remain binary, leading to the matrices $\mathbf{X}$ and $\mathbf{Y}$. These are the same as those in Figure 1 which we repeat here as

$$\mathbf{X}=\begin{bmatrix} 0&1&1&1&0\\ 1&0&1&0&1\\ 1&1&0&1&1\\ 1&0&1&0&1\\ 0&1&1&1&0 \end{bmatrix} \quad \text{and} \quad \mathbf{Y}=\begin{bmatrix} 0&1&1&1&1&0&0&0\\ 1&0&1&1&0&1&1&0\\ 1&1&0&0&0&0&1&1\\ 1&1&0&0&1&1&1&0\\ 1&0&0&1&0&1&0&1\\ 0&1&0&1&1&0&1&1\\ 0&1&1&1&0&1&0&1\\ 0&0&1&0&1&1&1&0 \end{bmatrix} \quad .$$

Distance: The Key Element in Explaining Movement Using Space Syntax and Spatial Interaction

The matrices $\mathbf{U}$ and $\mathbf{V}$ give the number of paths of one step between their respective nodes and $\mathbf{X}$ and $\mathbf{Y}$ show the one step paths excluding the self-loops. The simplest way of forming a distance between any two nodes in either the spatial interaction or the space syntax problem is to use a very well-known technique which involves computing these from powers of these matrices. As these operations are identical for any square matrix, we will only illustrate them for one of these – the $\mathbf{U}$ matrix – and simply state the related results for the other matrices. Then the number of paths of two step length between any two nodes is given by powering the number of one step paths in matrix $\mathbf{U}(1)=\mathbf{U}$ by $\mathbf{U}$, that is

$$\mathbf{U}(2)=\mathbf{U}(1)\mathbf{U}=\mathbf{U}^2 \qquad , \tag{4}$$

and by recursion, the number of n path lengths is

$$\mathbf{U}(n)=\mathbf{U}(n-1)\mathbf{U}=\mathbf{U}^n \qquad . \tag{5}$$

All the other results follow and we can state them as

$$\mathbf{V}(\ell)==\mathbf{V}^{\ell}; \quad \mathbf{X}(n)==\mathbf{X}^{n}; \quad \mathbf{Y}(\ell)==\mathbf{Y}^{\ell} \qquad . \tag{6}$$

These equations give the number of path lengths with and without the original self-loop and although we might conjecture that the number of path lengths co-varies with the accessibility or centrality of a node or street, then we still need to provide some measure of this accessibility from these lengths. In fact to generate all path lengths that are positive, then we need to power the matrices up to N or L to make sure we get all of these.

We can form two kinds of distance from these path lengths. First we argue that we wish to find the shortest path length for any pair of nodes and use this as a measure of distance. To do this we examine the number of path lengths at any iteration of equation (4) and if this number is different from zero when the number of path lengths on the previous iteration is zero, we set the distance to the value of the power – the step length. Then for $\mathbf{U}$

$$if\ U_{ik}(n) > 0\ and\ U_{ik}(n-1) = 0 \quad then \quad D_{ik}^{U} = n \tag{7}$$

and all the other step-length distances follow as $\mathbf{D}^V, \mathbf{D}^X, \mathbf{D}^Y$. This is a rather blunt measure in that it is consistent with binary step lengths but does not incorporate actual travel times or costs, Euclidean distances or travel costs which we will note a little later but for the time being this can be regarded as our first measure of accessibility or integration as it is sometimes called in space syntax. The second measure is based on a weighted sum of the path lengths. Let us assume a set of weights one for each step length and we call these w_z^U where $w_1^U > w_2^U > w_3^U > ...$ Note that we order these so closer/lower step lengths have more weight and these are then applied to the number of paths at each step. The second distance measure is thus

$$\delta_{ik}^{U}(n) = \sum_{z=1}^{n} w_z^U U_{ik}(z) \qquad , \tag{8}$$

and the other three measures can be defined accordingly $\delta_{ik}^X(n), \delta_{j\ell}^V(\ell), \delta_{j\ell}^Y(\ell)$.

It is now worth demonstrating what these two sets of distance measures actually show for our hypothetical example. The simplest distance measures are the step-distances where the value of the link between any two nodes is the number of steps a walker would have to make between one node and any other (for $\mathbf{X}$ via the street system) and between any street and any other (for $\mathbf{Y}$ via the intersection nodes). These matrices are easy to compute from the algorithm implied by equation (7) for our example. Then these step distances are

$$\mathbf{D}^X = [D_{ik}^X] = \begin{bmatrix} 2 & 1 & 1 & 1 & 2 \\ 1 & 2 & 1 & 2 & 1 \\ 1 & 1 & 2 & 1 & 1 \\ 1 & 2 & 1 & 2 & 1 \\ 2 & 1 & 1 & 1 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{D}^Y = [D_{j\ell}^Y] = \begin{bmatrix} 2 & 1 & 1 & 1 & 1 & 2 & 2 & 2 \\ 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 \\ 1 & 1 & 2 & 2 & 2 & 2 & 1 & 1 \\ 1 & 1 & 2 & 2 & 1 & 1 & 1 & 2 \\ 1 & 2 & 2 & 1 & 2 & 1 & 2 & 1 \\ 2 & 1 & 2 & 1 & 1 & 2 & 1 & 1 \\ 2 & 1 & 1 & 1 & 2 & 1 & 2 & 1 \\ 2 & 2 & 1 & 2 & 1 & 1 & 1 & 2 \end{bmatrix} ,$$

and it is easy to confirm that these simple path lengths with no more than 2 steps are those that result from casual inspection of the graphs in Figure 1.

These matrices are clearly very crude measures of accessibility but as they are very simple graphs and only based on binary relations, this is to be expected. The numbers of path computations are more detailed and these literally explode as we take more and more powers of the matrix for in bigger graphs, there is an exponentially growing number of circuits. To show this, we indicate the number of paths for the matrices $\mathbf{U}(N=5)=\mathbf{U}^5$ and for $\mathbf{V}(L=8)=\mathbf{V}^8$. These are computed as

$$\mathbf{U}^5 = \begin{bmatrix} 2019 & 1897 & 2929 & 1897 & 1776 \\ 1897 & 2019 & 2929 & 1776 & 1897 \\ 2929 & 2929 & 4660 & 2929 & 2929 \\ 1897 & 1776 & 2929 & 2019 & 1897 \\ 1776 & 1897 & 2929 & 1897 & 2019 \end{bmatrix} \quad \text{and}$$

$$\mathbf{V}^8 = \begin{bmatrix} 1246 & 1505 & 1165 & 1505 & 1165 & 1424 & 1424 & 1084 \\ 1505 & 1938 & 1505 & 1897 & 1424 & 1857 & 1897 & 1424 \\ 1165 & 1505 & 1246 & 1424 & 1084 & 1424 & 1505 & 1165 \\ 1505 & 1897 & 1424 & 1938 & 1505 & 1897 & 1857 & 1424 \\ 1165 & 1424 & 1084 & 1505 & 1246 & 1505 & 1424 & 1165 \\ 1424 & 1857 & 1424 & 1897 & 1505 & 1938 & 1897 & 1505 \\ 1424 & 1897 & 1505 & 1857 & 1424 & 1897 & 1938 & 1505 \\ 1084 & 1424 & 1165 & 1424 & 1165 & 1505 & 1505 & 1246 \end{bmatrix} .$$

The weights for combining the path numbers up to the total number of steps $\mathbf{U}(N=5)=\mathbf{U}^5$ and $\mathbf{V}(L=8)=\mathbf{V}^8$ are based on the simple expediency of making the weight proportional to the total maximum path lengths N or L less the step length being considered, that is, $w_i^U(n) \propto N-n+1$ or $w_i^V(\ell) \propto L-\ell+1$. The weighted distances for the spatial interaction and space syntax variants are thus computed as

$$\boldsymbol{\delta}^U(5) = [\delta_{ik}^U(5)] = \begin{bmatrix} 194 & 176 & 269 & 176 & 159 \\ 176 & 194 & 269 & 159 & 176 \\ 269 & 269 & 437 & 269 & 269 \\ 176 & 159 & 269 & 194 & 176 \\ 159 & 176 & 269 & 176 & 194 \end{bmatrix} \quad \text{and}$$

$$\boldsymbol{\delta}^V(8) = [\delta_{ik}^V(8)] = \begin{bmatrix} 12730 & 16167 & 12594 & 16167 & 12594 & 16030 & 16030 & 16030 \\ 16167 & 20712 & 16167 & 20643 & 16030 & 20575 & 20643 & 20643 \\ 12594 & 16167 & 12730 & 16030 & 12457 & 16030 & 16167 & 16167 \\ 16167 & 20643 & 16030 & 20712 & 16167 & 20643 & 20575 & 20575 \\ 12594 & 16030 & 12457 & 16167 & 12730 & 16167 & 16030 & 16030 \\ 16030 & 20575 & 16030 & 20643 & 16167 & 20712 & 20643 & 20643 \\ 16030 & 20643 & 16167 & 20575 & 16030 & 20643 & 20712 & 20712 \\ 12457 & 16030 & 12594 & 16030 & 12594 & 16167 & 16167 & 16167 \end{bmatrix} .$$

We are at last in a position to say something about these distances/path numbers/step lengths relative to the problem shown in Figure 1. It is very clear that there is little discrimination between the relative positioning of the nodes as intersections in the primal and the nodes as streets in the dual. That is, the graphs are symmetric and strongly connected and intuitively if we were to measure the relative importance of the nodes in each graph, their in-degrees (and out-degrees as the graphs are symmetric) would not show much variation. In the primal it is clear that the central node 3 seems most important while in the dual, nodes 1, 4, 6 and 7 form a central block that has more importance than the outer block that consists of nodes 1, 3, 5 and 8.

Spatial interaction models usually predict both the relative flows between nodes that take place along streets as well as the total flows destined for each origin and destination (nodes) while space syntax is associated with flows along streets that need to be aggregated from the relative positioning of any one street connected to all others. In short, both models make use of accessibilities which are summations of interactions; in spatial interaction, we predict flows from information about $i \leftrightarrow k$ nodes as well as flows into i and k whereas in space syntax, the flows between streets $j \leftrightarrow \ell$ have no meaning and what we need to do is model the notional flows that take place on each street j and ℓ. Thus it is accessibilities that we need to be concerned with here. As these are all defined the same way as summations of distance measures into nodes whether these nodes be intersections or streets, then we will just illustrate these for the step distances $\mathbf{D}^U$ and $\mathbf{D}^V$. As space syntax uses the step distance measures $\mathbf{D}^V$ as the core element in its tool box, we first define the total step distance for street j as d_j^V and normalise this by the maximum step distance m as $\hat{d}_j^V$ so that comparisons can be made between systems with different numbers of streets; these measures are

$$\left.\begin{aligned} d_j^V &= \sum_\ell D_{j\ell}^V \\ \hat{d}_j^V &= \frac{d_j^V}{m} = \frac{\sum_\ell D_{j\ell}^V}{m} \end{aligned}\right\} \quad . \tag{9}$$

These measures although referred to as measures of integration by Hillier and Hanson (1985) are in fact measures of ‘inaccessibility’ and in most space syntax applications, the inverse of this measure is used to define what they call integration. In most recent applications, the measure integration appears to have been dropped and the more common measure of accessibility following conventional usage in spatial interaction and transportation modelling after Hansen (1959) is now being used. In fact the measure is usually taken as the inverse of depth d_j^V or $\hat{d}_j^V$ and normalised to sum to 1, that is

$$\bar{d}_j^V = \frac{1/d_j^V}{\sum_\ell 1/d_\ell^V} , \qquad \sum_j \bar{d}_j^V = 1 \quad . \tag{10}$$

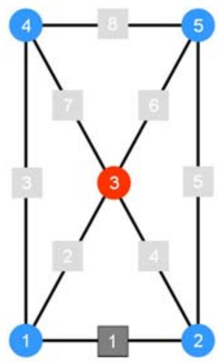
This measure is sometimes referred to as real relative asymmetry although it is unclear where the term comes from and what the asymmetry is that is being referred to (Bafna, 2003). What happens in space syntax is that as this measure in equation (10) is associated with the street system, the relative variations in the measure (which is the average or total depth of any one street to all others) is plotted for each street across the red-yellow-green-blue colour spectrum to produce the typical space syntax map. The primal problem has accessibilities associated with the nodes which are locations at street intersections. In exactly the same way, we form the same accessibilities and for completeness we define these as follows

$$d_i^U = \sum_k D_{ik}^U \; ; \quad \hat{d}_i^U = \frac{d_k^U}{n} = \frac{\sum_k D_{ik}^U}{n} \; ; \quad \bar{d}_i^U = \frac{1/d_k^U}{\sum_k 1/d_k^U} \quad , \quad \sum_i \bar{d}_i^U = 1 \quad . \qquad (11)$$

The accessibilities show the relative intensity of flows at an intersection. If it is required to examine particular flows as in the space syntax dual, then the actual distance measures need to be used to make these comparisons but this has not been done in space syntax for in computing these distance matrices, few have broached the kind of predictive modelling that spatial interaction requires. In short although space syntax focuses primarily on the geometry and topology of the street network, the street network is simply the starting point for spatial interaction and the accessibility measures – which are in fact an essential part of spatial interaction modelling – are used quite differently from those in space syntax.

At this point, we have a common framework from computing relative measures of nearness or accessibility in both the primal and the dual and in Figure 2 we show how we can map these to show variations in intensity for both the distance-step matrices and their accessibilities. We will not explore the weighted path number distances in equations (8) at this point here but keep these in mind for later applications.

$$d_i^U = \sum_k D_{ik}^U = \begin{bmatrix} 7 \\ 7 \\ 6 \\ 7 \\ 7 \end{bmatrix} \qquad \hat{d}_i^U = \frac{d_k^U}{5} = \begin{bmatrix} 1.4 \\ 1.4 \\ 1.2 \\ 1.4 \\ 1.4 \end{bmatrix} \qquad \bar{d}_i^U = 100\frac{1/d_k^U}{\sum_k 1/d_k^U} = \begin{bmatrix} 19.355 \\ 19.355 \\ 22.581 \\ 19.355 \\ 19.355 \end{bmatrix}$$



$$d_j^V = \sum D_{j\ell}^V = \begin{bmatrix} 12 \\ 11 \\ 12 \\ 11 \\ 12 \\ 11 \\ 11 \\ 12 \end{bmatrix} \qquad \hat{d}_j^V = \frac{d_j^V}{8} = \begin{bmatrix} 1.5 \\ 1.375 \\ 1.5 \\ 1.375 \\ 1.5 \\ 1.375 \\ 1.375 \\ 1.5 \end{bmatrix} \qquad \bar{d}_j^V = 100\frac{1/d_j^V}{\sum_\ell 1/d_\ell^V} = \begin{bmatrix} 11.957 \\ 13.043 \\ 11.957 \\ 13.043 \\ 11.957 \\ 13.043 \\ 13.043 \\ 11.957 \end{bmatrix}$$

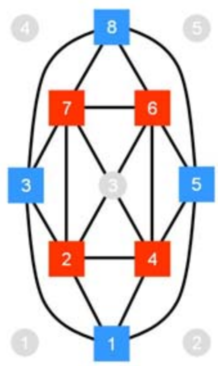


Figure 2: The Primal and Dual Spatial Graphs and Their Accessibilities
The Nodes in the Primal are Street Intersections and the Nodes in the Dual are the Streets

You can see clear relations between the primal – the spatial interaction problem – and the dual – the space syntax problem – where we simply map the accessibilities into street intersections and street segments in the primal and dual respectively but on the planar graph of the network which we show in Figure 3. It is essential to note that there is an intrinsic asymmetry between spatial interaction and space syntax in that we use the planar graph which lies at the basis of the network in spatial interaction to represent both problems. In short, in space syntax we collapse the movement onto nodes that define the streets whereas in spatial interaction we deal directly with movement on streets as we will elaborate below.

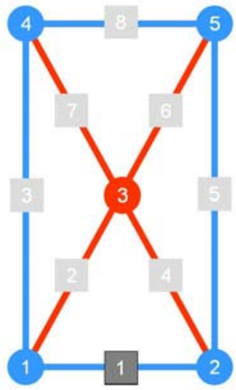


Figure 3: Accessibility Levels for the Primal and Dual Problems on the Planar Network

Models for Predicting Movement and Location

Operations on the primal and dual network graphs do not lead to predictions of movement but to measures of connectivity which define indices of accessibility and distance. These indices might then be considered as being variables that can be associated or compared with activities at locations or movement along streets between locations but this is an additional stage in the analysis. In spatial interaction modelling for example, distances are central to predicting movements but these are usually defined *a priori* and although measures such as the step distance (or the path) matrices $\mathbf{D}^U, \mathbf{D}^X$ and their accessibilities $\mathbf{d}^U, \mathbf{d}^X$ could be used as independent variable inputs, they are not quite in the form required for the standard models. In terms of the distance matrices $\mathbf{D}^V, \mathbf{D}^Y$ and their accessibilities $\mathbf{d}^V, \mathbf{d}^Y$ for the dual space syntax problem, these are the only elements that can be used to predict movement, and in this case, the distance matrices simply indicate notional flows or interactions between streets, that is $D_{j\ell}^V$ which do not have the meaning of actual flows of traffic *per se*. Only when these are aggregated to d_j^V are we able to compare these to the flow on the relevant street segments $i \leftrightarrow k$ in order to see how good the fit is to real data.

There are two problems with doing this however. First using the step lengths, the range of step lengths whereby a network becomes completely connected might be very narrow. If you look at Figure 1(b), the dual graph, then as we have already worked out, there are only two step lengths – 1, 2 – before the dual is completely connected; this is far too small a variation to use in computing accessibilities even though the range widens once the number of nodes is increased. In fact in a large network in the form of a chain, then the range of accessibilities would vary as the number of nodes. However in typical networks which tend to be at best a small set of large monocentric clusters at whatever scale of town one is looking at, the range is much narrower than we might expect would explain variations in movement. This is a major problem in space syntax and is seen in the fact that when comparisons of accessibility measures from the dual are made with movements along street segments, the scatter graphs are characterised by a small number of measures of accessibility all at integer value, and a much larger number of measures of movement. The appearance of these graphs in fact is not a random scatter but a more structured striation.

The second problem is that the accessibilities which assume that a street is a node in the dual, are derived from links to other streets – other nodes in the dual – and that this implies some sort of flow from these other streets. In short it is not clear that if we aggregate, say, the step distances in the dual to get the accessibility of a street as $d_j^V = \sum_\ell D_{j\ell}^V$, then the interactions between streets $D_{j\ell}^V$ are actually flows. In spatial interaction modelling, however the flows are always unambiguously associated with movements as measured by vehicular passenger traffic, migration, freight and so on. There is a third problem that is more generic. In spatial interaction modelling, flows are predicted between all intersections or nodes in the street network whereas in space syntax, the underlying planar graph does not connect everywhere with everywhere else directly and thus flows can only be measured on the direct links in the graph. In short whereas in spatial interaction modelling as we have direct links such as ik and kz, we also have iz which does not necessarily exist as a line segment in the planar street graph. In short, in space syntax, we only examine direct links in the graph which are associated with nodes j and thus many possible links do not appear in the graph whereas the implicit graph in spatial interaction modelling is completely connected. In short, in spatial interaction modelling, we distribute trips to all possible links between intersections or nodes regardless of whether a separate physical links exists whereas in space syntax, the flows are implicitly associated with those on a segment that are measures of traffic. Spatial interaction distributes trips whereas space syntax assumes these trips have been already assigned to a physical network based on direct street segments.

In fact what has been done in space syntax is to construct models that explain movement as function of the direct street segments in the graph using street accessibility. Defining the observed movement in a street as T_{ik}^{obs}, we assume a simple regression such as

$$T_{ik}^{obs} = \alpha + \beta d_j^V \text{ where } j \text{ is the same as } i \leftrightarrow k \quad . \qquad (12)$$

Hillier, Penn, Hanson, Grajewski, and Xu (1993) refer to this relationship as that governing ‘natural movement’ and their work shows that the only flows that are compared with accessibility are those that are measured as composite totals on each link. These are not broken down into flows between all nodes in the street graph, and thus implicitly occur after spatial interactions have been assigned to network links. This paper also reveals the problem of striation referred to above which concerns the fact that the accessibility values are integers and cover a narrow range. The levels of variance explained associated with these kinds of regressions are rarely more than 0.6 and due to the nature of the data and very often the small number of distinct observations, this would not be regarded as a satisfactory predictive model. In my view, the advantages of space syntax lie elsewhere in much more qualitative but structured discussion of how space is formed and how it is moulded with respect to generic human interactions. In this context, it plays an important role but essentially as we implied at the outset, it is not a predictive model.

Notwithstanding this rather negative consequence, we will argue below that as it is virtually impossible to take the standard spatial interaction model and derive space syntax from this and vice versa, it is still possible to make progress by making changes to the formulation of both spatial interaction and space syntax and casting these in a form where direct comparison and derivations of one from the other can be made. However before this, we need to introduce spatial interaction as a predictive model because it is still possible to use measures from the primal to structure its predictive capabilities. The clearest way of introducing one of the many variants of spatial interaction is in conditional probability terms. Then

$$\left.\begin{aligned} T_{ik} &= E_i p_{k|i} \\ E_i &= \sum_k T_{ik} = E_i \sum_k p_{k|i} \end{aligned}\right\} \quad \text{where} \quad \sum_k p_{k|i} = 1 \qquad . \tag{13}$$

T_{ik} is the flow or trips from zone or intersection i to k, E_i is some measure of the size of activity at location/intersection i which is the flow to be distributed as trip interactions, and $p_{k|i}$ is the probability or fraction of E_i which is distributed as trips to k. This probability model is usually configured as the product of an attractor of the zone k, F_k and some function of the generalised distance/travel cost c_{ik} from i to k which we hypothesise as

$$p_{k|i} = \frac{F_k \exp(-\lambda c_{ik})}{\sum_z F_z \exp(-\lambda c_{iz})} \qquad . \tag{14}$$

The independent variables are F_k and c_{ik} and the model is calibrated by finding the value of the parameter λ which minimises some statistic of difference between observed and predicted trips $g(T_{ik}^{obs} - T_{ik})$. In terms of the operations on the primal network graph, then it is easy to see that c_{ik} could be one of the measures derived earlier so in this form, additional variables are defined, or at least there needs to be a driver for trip-making or movement such as E_i. There are many variants of these models and the model in this form is called singly-or origin-constrained (Wilson, 1970).

In fact we might use the accessibility values for the dual in equation (14), rather than the distance values in the primal and were we to drop the attractor, then the equation for the spatial interaction model becomes

$$p_{k|i} = \frac{\exp(-\lambda d_j^V)}{\sum_z \exp(-\lambda d_z^V)} \quad \text{where } j \text{ is the same as } i \leftrightarrow k \qquad , \tag{15}$$

and this can be calibrated in the same way. This method of coupling spatial interaction to space syntax through the widely used measure of integration (accessibility) shows one way of integrating the two models but due to the ambiguities about this index, we consider this to be a weak method. Essentially spatial interaction relies much more on Euclidean distance as some function of the generalized cost of travel. Nevertheless were we to use the number of paths from the space syntax problem $\mathbf{V}$ or $\mathbf{Y}$ and the accessibilities formed from these, weighted over many step lengths or simply based on some high step length. This might be a preferable variant following equation (15) which we can write as

$$p_{k|i} = \frac{\exp(-\lambda \sum_{\ell} V_{j\ell})}{\sum_{z} \exp(-\lambda \sum_{\ell} V_{z\ell})} \quad \text{where } j \text{ is the same as } i \leftrightarrow k \quad . \tag{16}$$

This is closer to the model in equations (13) and (14) where we use generalised travel cost which incorporates some measure of distance but it is still a very weak coupling and is unlikely to find favour with those who consider that much more powerful functions of deterrence need to be used.

It is most unlikely that we can do better than this at this stage for what we need is a much stronger method of integration. We already have the key to this for it resides in the coupling of the bipartite graph whose matrix is $\mathbf{A}$ which separates nodes from lines in the original planar street network. At this point, let us speculate that the way forward lies in this approach and what we need to do is find much better measures of distance that take account of this coupling other than those based on step lengths. The method we will adopt has been used before in several contexts by the author (Batty, 2013) and it consists in slightly changing the nature of the two models so that they intersect in a much more basic way. This we will broach in the next section before we produce an integrated model, ultimately demonstrating this on a large but simplified network of links and zones in Greater London.

A Probabilistic Interpretation of Distance and Connectivity

We can first convert the raw interaction matrices $\mathbf{U}$ and $\mathbf{V}$ into stochastic matrices where we interpret the cells as being the probability of a node relating to another node and the probability of a street relating to another street respectively. Then we define these probabilities as

$$P_{ik} = \frac{U_{ik}}{\sum_{k} U_{ik}} \quad , \quad \sum_{k} P_{ik} = 1 \quad , \quad \text{and} \tag{17}$$

$$Q_{j\ell} = \frac{U_{j\ell}}{\sum_{\ell} U_{j\ell}} \quad , \quad \sum_{\ell} Q_{j\ell} = 1 \quad . \tag{18}$$

We can interpret these as follows. If walker starts at a node i, and then with probability P_{ik} moves to node k, then the probability of that same walker visiting a node z on the next step is given as $\sum_{k} P_{ik} P_{kz}$ which is the respective element of the second power of the matrix. Then on the n'th step of the walk, we can compute the probability as $P_{iz}(n) = \sum_{k} P_{ik}(n-1) P_{kz}$ which in matrix terms is given as $\mathbf{P}^{n} = \mathbf{P}^{n-1}\mathbf{P}$. This sequence defines a discrete Markov chain and if the matrix $\mathbf{P}$ is strongly connected which it must be for the problem to be meaningful and the street system connected, then it is well known that the limit of this sequence is a fixed point vector which we can call $\mathbf{p}$. In short, if we begin the walk with a probability vector

$\mathbf{p}(1)$, the sequence updating this vector can be written as $\mathbf{p}(n) = \mathbf{p}(n-1)\mathbf{P} = \mathbf{p}(1)\mathbf{P}^{n-1}$. As this vector converges to $\mathbf{p}$, then in the limit we can solve for $\mathbf{p}$ from $\mathbf{p} = \mathbf{pP}$.

An exactly analogous process exists if we begin with a walker on a street j who visits another street ℓ with probability $Q_{j\ell}$. In the limit, we can solve for the steady state probability vector $\mathbf{q} = \mathbf{qQ}$ and this gives the overall probability of any walker visiting a street if the walk continues indefinitely. As in all Markov chains, the initial distribution of probabilities washes out and this is something that we are not sure is desirable for it implies that the initial structure exerts a decreasing effect on the final state. In this steady state, a walker has the same chance as any other of visiting a node or street regardless of where they started from. The key issue is what these vectors actually imply. In fact, they are measures of the number of walkers travelling to different places in the system; as such, these may correlate with any of the distance measures introduced earlier and we will test this correlation in another worked example below. For the moment, let us simply note that $\mathbf{p}$ and $\mathbf{q}$ are accessibility vectors for intersections and streets defined from the random walks associated with the different probability processes $\mathbf{P}$ and $\mathbf{Q}$. We can in fact define associated processes based on the sliced data matrices $\mathbf{X}$ and $\mathbf{Y}$ but we will not do so here as these are close to processes defined on the raw interaction data. One last point before we move to a deeper view of these processes: clearly as $\mathbf{U} = \mathbf{AA}^T$ and $\mathbf{V} = \mathbf{A}^T\mathbf{A}$, then the primal and dual processes $\mathbf{P}$ and $\mathbf{Q}$ are related. We can write these as $\mathbf{P} = \boldsymbol{\delta}^p\mathbf{AA}^T$ and $\mathbf{Q} = \boldsymbol{\delta}^q\mathbf{A}^T\mathbf{A}$ where $\boldsymbol{\delta}^p$ and $\boldsymbol{\delta}^q$ are diagonal matrices defined to ensure that $\mathbf{P}$ and $\mathbf{Q}$ are row stochastic. Some manipulation of these relations suggests that there are more explicit links between their steady state vectors in terms of the initial matrices $\mathbf{A}$ and $\mathbf{A}^T$ but we have not taken this further as yet. Our purpose here is to work with relations where we define the probabilities at a more elemental level.

To introduce these, we can define the probability structure that determines distance measures on the underlying graph in terms of the basic data matrix $\mathbf{A}$. Let us define the probability of a node belonging to a street as

$$G_{ij} = \frac{A_{ij}}{\sum_j A_{ij}} \quad , \quad \sum_j G_{ij} = 1 \quad , \qquad \text{and} \tag{19}$$

a street belonging to a node as

$$C_{jk} = \frac{A^T_{jk}}{\sum_k A^T_{jk}} \quad , \quad \sum_k C_{jk} = 1 \qquad . \tag{20}$$

Now these matrices are stochastic and by concatenating them to form the primal and dual probabilities, we define

$$\hat{P}_{ik} = \sum_j G_{ij}C_{jk} \quad , \quad \sum_k \hat{P}_{ik} = 1 \qquad , \qquad \text{and} \tag{21}$$

$$\hat{Q}_{j\ell} = \sum_k C_{jk}G_{k\ell} \quad , \quad \sum_\ell \hat{Q}_{j\ell} = 1 \qquad . \tag{22}$$

These matrices have the following interpretation which are further measures of distance. The matrix $\hat{\mathbf{P}}$ records the probability of a walker at node i accessing another node k via any street j while the matrix $\hat{\mathbf{Q}}$ gives the probability of moving from a street j to another street ℓ via any node k. To detail this, a walker at any node i has a probability of accessing each street j and from each of these streets, he/she has a probability of reaching another node k. The same type of fixed point vectors result from this process of continually moving from node to node or street to street as we indicated above for the processes based on $\mathbf{P}$ and $\mathbf{Q}$. This washes out the initial urban structure and insofar as we can define the resultant probabilities in the steady state as accessibilities, these are defined from $\hat{\mathbf{p}} = \hat{\mathbf{p}}\hat{\mathbf{P}}$ and $\hat{\mathbf{q}} = \hat{\mathbf{q}}\hat{\mathbf{Q}}$.

These primal and dual processes hold the key to the integration between spatial interaction through the primal and space syntax through the dual. Let us write the steady state equation for intersection nodes as $\hat{\mathbf{p}} = \hat{\mathbf{p}}\hat{\mathbf{P}} = \hat{\mathbf{p}}\mathbf{GC}$. Now if we post-multiply this by $\mathbf{G}$, we get $\hat{\mathbf{p}}\mathbf{G} = \hat{\mathbf{p}}\mathbf{GCG} = \hat{\mathbf{p}}\mathbf{G}\hat{\mathbf{Q}}$. Now as the steady state vector $\hat{\mathbf{q}}$ associated with $\hat{\mathbf{Q}}$ is unique, it is clear that $\hat{\mathbf{p}}\mathbf{G} = \hat{\mathbf{p}}\mathbf{G}\hat{\mathbf{Q}} = \hat{\mathbf{q}} = \hat{\mathbf{q}}\hat{\mathbf{Q}}$. Thus it is clear that $\hat{\mathbf{q}} = \hat{\mathbf{p}}\mathbf{G}$ and in like manner, $\hat{\mathbf{p}} = \hat{\mathbf{q}}\mathbf{C}$. This is a very clear relation between the two processes and it is the simplest way they interlock. What they mean is as follows: writing the steady state relations in full as

$$\hat{p}_k = \sum_j \hat{q}_j C_{jk} \quad \text{and} \quad \hat{q}_j = \sum_i \hat{p}_i G_{ij} \quad ,$$

then if you are at a node k, then the probability of being there is equal to the probability of being on any street which is connected to that node, while if you are on a street, the probability of being there is equal to the probability of being at any node that is associated with that street.

A Comparison of Distance and Probability Measures

We are now in a position to compare all these measures and to this end, we will introduce a second more structured graph so that we are able to develop some intuition about the relatively positioning of streets and their intersections/nodes. This is shown in Figure 4(a) where it is clear that nodes 6 and 9 and possibly 4 are the most central and connected while streets 6 and 9, then 7 and 8 seem the most accessible, although this is harder to guess from the configuration. However this is to be tested below using the various accessibility measures. It is now worth stating the distance measures that we will compute from all those introduced in the previous sections. We will list these, noting that for many of these measures, these are identical when defined for either origin or destination nodes or streets due to the fact that the interaction matrices are symmetric. We will define the measures for the primal and dual on the same line below and annotate them with respect to their meaning:

1. basic in-degrees $\quad u_i = \sum_k U_{ik} \quad v_j = \sum_\ell V_{j\ell}$
2. sliced in-degrees $\quad x_i = \sum_k X_{ik} \quad y_j = \sum_\ell Y_{j\ell}$
3. weighted paths $\quad \delta_i^U = \sum_k \delta_{ik}^U \quad \delta_j^V = \sum_\ell \delta_{j\ell}^V$

4. sliced weighted paths $\delta_i^X = \sum_k \delta_{ik}^X$ $\delta_j^Y = \sum_\ell \delta_{j\ell}^Y$
5. inverse step lengths $\Delta_i^U = 1/\sum_k D_{ik}^U$ $\Delta_j^V = 1/\sum_\ell D_{j\ell}^V$
6. sliced inverse step lengths $\Delta_i^X = 1/\sum_k D_{ik}^X$ $\Delta_j^Y = 1/\sum_\ell D_{j\ell}^Y$
7. aggregate probabilities p_i q_j
8. disaggregate probabilities $\hat{p}_i$ $\hat{q}_j$

There is one last measure that we will introduce. So far none of our measures incorporate any measure of Euclidean distance. In fact for each street segment j, we can define a measure of distance of generalised travel cost d_j. We now augment our raw data matrices by weighting each node-street link ij by d_j and we now form a new raw matrix (and its transpose follows directly from this) as $\overline{A}_{ij} = A_{ij}\exp(-\lambda d_j)$. We use these matrices to construct new values for the matrices $\mathbf{G}$ and $\mathbf{C}$ and from this, we compute new steady states which we call $\overline{\mathbf{p}}$ and $\overline{\mathbf{q}}$. These form our ninth measure:

9. weighted probabilities $\overline{p}_i$ $\overline{q}_j$

In this formulation, we now have a parameter λ which we can use to moderate the effect of distance. Moreover for any of the limit probabilities we can take the probabilities that pertain to any power z of the matrices in question (for measures 7-9) and also use this as a parameter; that is, choose the relevant probabilities that optimise the fit of the model to data but more of this later when we come to empirical applications.

We have computed all these measures for the graph in Figure 4(a) and to compare them, we will correlate them. So for the primal problem measures, we show these correlations in Table 1(a) and for the dual in Table 1(b).

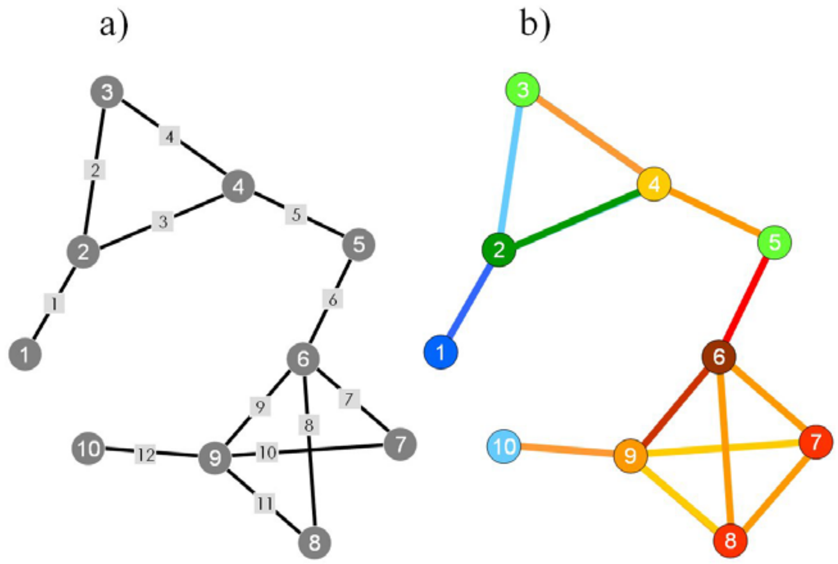


Figure 4: The Planar Graph/Street Network: a) Nodes and Arcs Labelled b) Nodes and Arcs Coloured According to Inverse Step (Accessibility) Values

a)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	**1.00**								
2	0.97	**1.00**							
3	0.86	0.72	**1.00**						
4	0.85	0.72	0.99	**1.00**					
5	0.71	0.76	0.52	0.58	**1.00**				
6	0.90	0.89	0.73	0.76	0.80	**1.00**			
7	1.00	0.97	0.86	0.85	0.71	0.90	**1.00**		
8	0.97	1.00	0.72	0.72	0.76	0.89	0.97	**1.00**	
9	0.74	0.63	0.81	0.80	0.54	0.70	0.74	0.63	**1.00**

b)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	**1.00**								
2	0.93	**1.00**							
3	0.89	0.85	**1.00**						
4	0.72	0.76	0.95	**1.00**					
5	0.62	0.72	0.57	0.59	**1.00**				
6	0.77	0.85	0.72	0.69	0.96	**1.00**			
7	1.00	0.93	0.89	0.72	0.62	0.77	**1.00**		
8	0.92	0.75	0.70	0.48	0.51	0.63	0.92	**1.00**	
9	0.08	0.13	0.24	0.30	0.11	0.19	0.08	-0.11	**1.00**

Table 1: Correlations Between Selected Distance Measures for the a) Primal Spatial Interaction and b) Dual Space Syntax Problems

In both problems, the measure which correlates most with all 8 other measures is in fact the in-degree for the basic matrices $\mathbf{U}$ and $\mathbf{V}$. However, we have chosen the inverse step length from the sliced data matrices $\mathbf{X}$ and $\mathbf{Y}$ and we have plotted the values of these in Figure 4(b). Our initial intuition on the relative importance of the nodes is clearly born out with node 6 occupying a pivotal position and 7 and 8 next with 9 close behind. The less connected and more extreme nodes such as 1 and 10 are the least important. Although this is not a definitive demonstration of the relatedness between nodes and streets, the space syntax measures show that streets 6 and 9 are the most important with 4, 5, 7 and 8 being the next important. These mirror the nodal structure also shown in Figure 4(b). It is worth noting that the last measures – the ninth based on $\overline{\mathbf{p}}$ and $\overline{\mathbf{q}}$ – have been defined using random distances, that is $d_j = rand(100)$, and it is no surprise that the lowest correlations with the other measures occur here. In fact the steady states of the probability measures 7 and 8 are quite highly correlated with the other measures but note that for measure 7, this has complete correlation with the in-degree measure 1 for both the primal and duals while measure 8 has a complete correlation with the sliced in-degree with the primal. In fact what is clear from this is that if we have a very simple structure where the number of in-degrees is the same for each node, then the probability measures are likely to have a very high correlation with these in-degrees (Batty, 2016). This can be very problematic where we control the nodes or where we limit the connections. It means that the less we differentiate urban structure through connectivity, the less differences there are between the measures of accessibility, a problem that is quite significant in the empirical applications that now follow.

An Empirical Demonstration of Primal-Dual Integration

As we have emphasised, space syntax and spatial interaction represent space at different scales with space syntax dealing with the literal physical connections between places while spatial interaction deals with generic movements between origins and destinations which can then be assigned to the finer scale physical network that space syntax takes as its starting point. Spatial interaction deals with all flows between origins and destinations of magnitude N^2 whereas space syntax deals with a subset of these flows $L << N^2$ where only those on the physical links of the network are considered. However as we have already illustrated, to compare the two approaches, we need to begin with a common network and to this end we have constructed a physical network of links from the generalised travel cost and distances between some 633 zones which comprise the Greater London Authority (GLA) area. These zones are based on wards, the most basic electoral districts which are associated with the 33 boroughs that make up the area, with on average these zones containing 11,330 resident population and 7,181 employment. We show the zones and their centroids in Figure 5(a) where we indicate the GLA area in its wider zonal context.

The level of detail of the street network is well below this scale so in this application, what we will do is construct a synthetic network from the distance links that are used to form the generalised travel cost matrix $[c_{ik}]$. To build this network, we first take the 5 shortest links for each zone i and in cases where this does not lead to symmetric links, that is where we have a shortest links c_{ik} but do not have a link c_{ki}, we add this link to the network, thus giving us at least 5 links from every origin to its nearest neighbour destinations. We also need to ensure that on the edge of the area, we also take 5 such links and thus we have a ring of centroids in zones outside the GLA area which connect to the 633 zones inside the area. This increases the number of zones to 699, with some 66 acting as edge zones outside the area. In total, we define 3944 links from these 699 zones giving an average in-degree (and out-degree) of 5.64, a little greater than the 5 chosen initially for each zone. The total number of links is a small percentage of the total possible links, the ratio being less than 1% $(0.008 = 3944/N^2 = -3944/488{,}601)$ which is an extremely sparse matrix. It is arguable as to whether or not this network is sufficiently rich to pick up the urban structure and connectivity of London but at least the links chosen do exist between largely between adjacent zones. In this sense, the network can be referred to as a ‘nearest neighbour’ network and we show its form in Figure 5(b).

From the previous hypothetical example, the sliced in-degree measures 1 and 2 are the most highly correlated with all others and we consider these to be a natural baseline for planar graphs in terms of their direct accessibilities. Moreover it is very clear that these measures pick up local structure although all the other measures are based on indirect as well as direct links some with an appropriate weighting. We show the sliced in-degrees for nodes $x_i = \sum_k X_{ik}$ and intersections $y_j = \sum_\ell Y_{j\ell}$ which reflect the primal and dual problems respectively in Figure 6(a) and (b) and it is immediately obvious that these measures reflect the fact that the network has been constructed using a rule of thumb starting with 5 links per node. This is likely to give a much more muted distribution of accessibilities for nodes and well as paths and this is in fact the case in Figure 6. Moreover the local structure is picked up very clearly in street accessibilities in the dual in Figure 6(b) while the accessibility in the nodes is dominated by a handful of local nodes that because of their physical juxtaposition come out as being more central than the others. It is worth noting that the in-degree structure based on $\mathbf{x}$ and $\mathbf{y}$ are equivalent to the steady state vectors from the aggregate steady state probability vectors $\hat{\mathbf{p}} = \hat{\mathbf{p}}\hat{\mathbf{P}}$ and $\hat{\mathbf{q}} = \hat{\mathbf{q}}\hat{\mathbf{Q}}$.

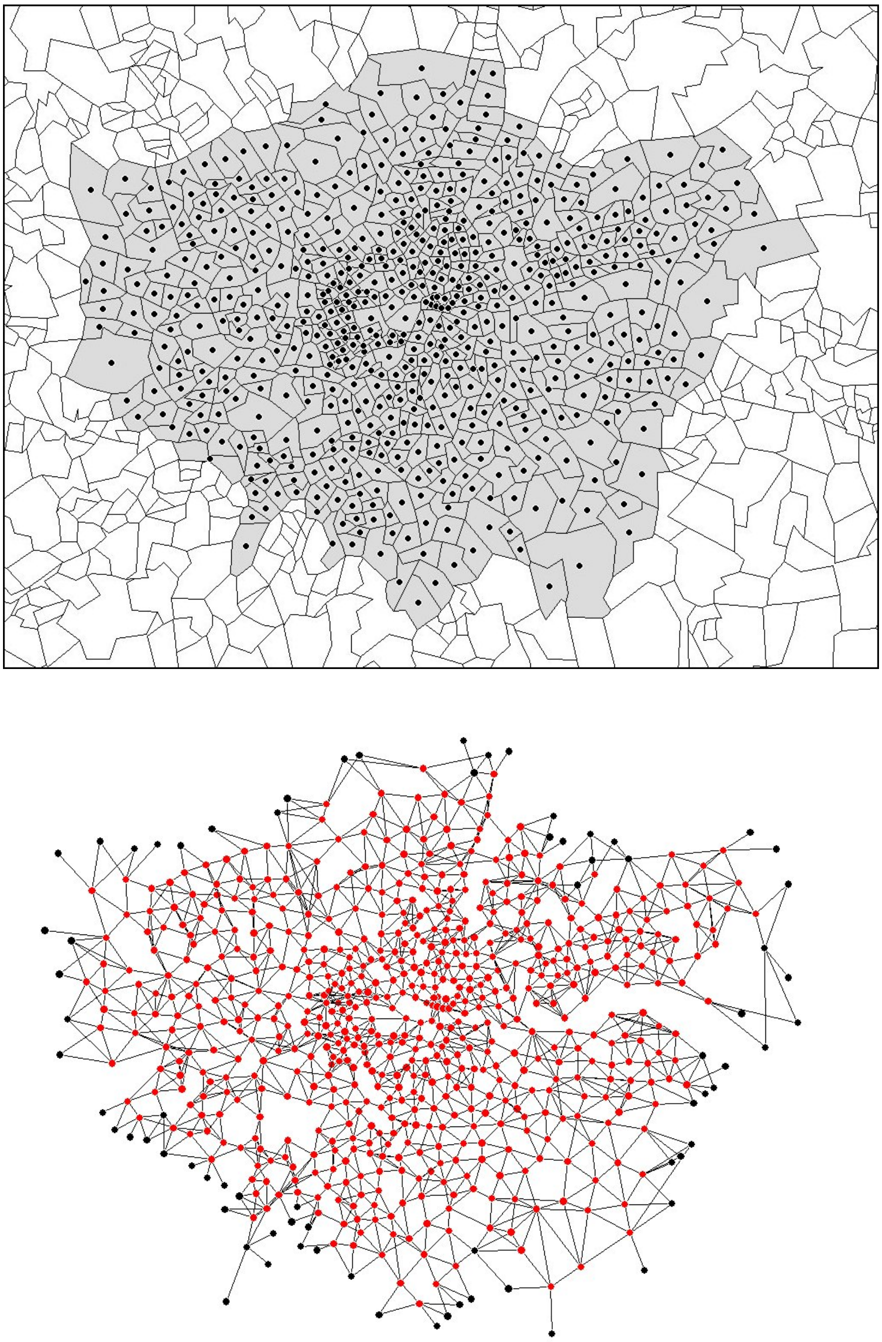


Figure 5: The Zoning System, Centroids, and Network Links for the London Area

(a) The zones in the GLA Area are coloured grey while in (b) the nodes external to the area but within the network are coloured black

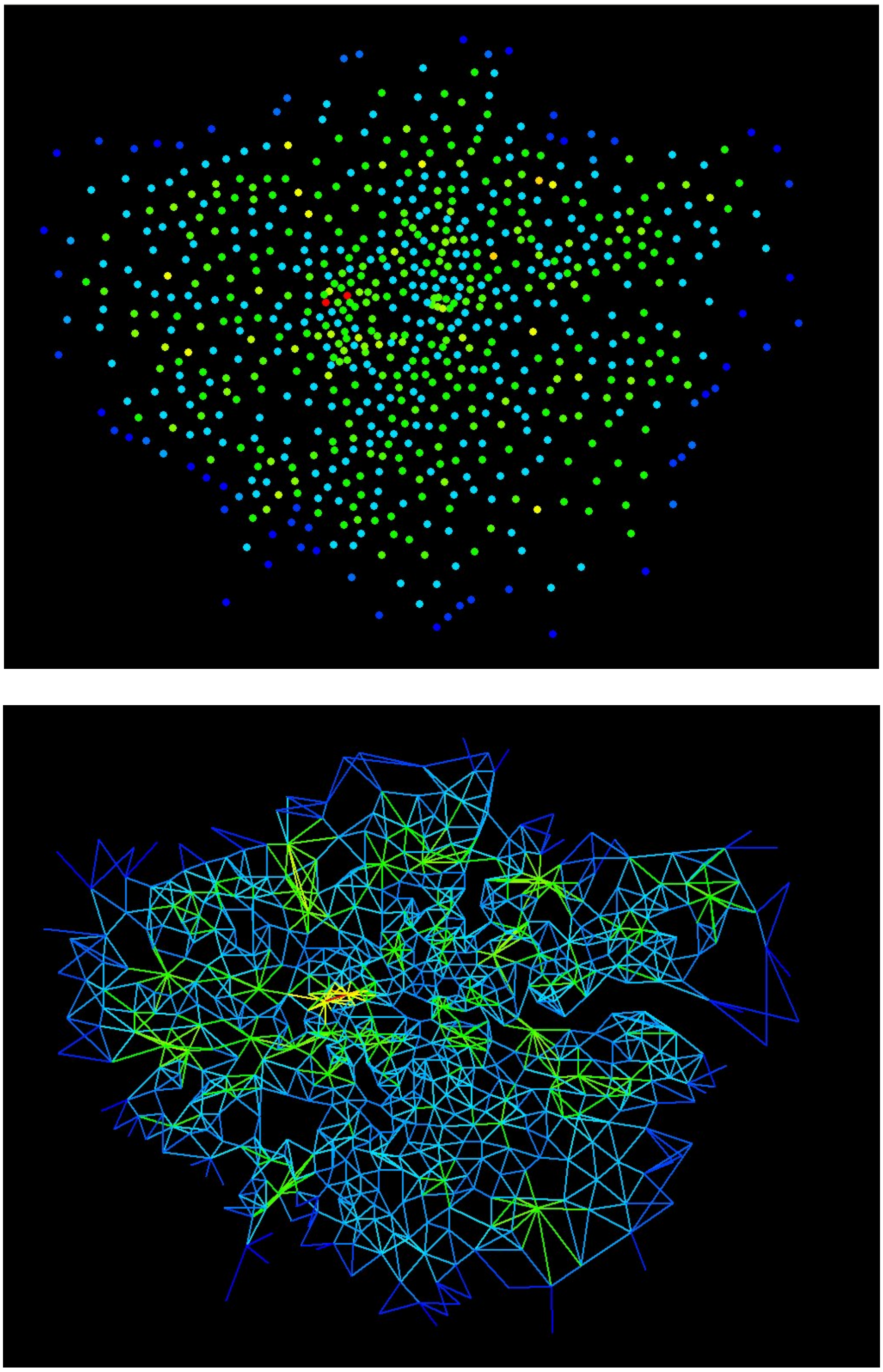


Figure 6: Nodes (a) and Street Accessibilities (b) Based on In-degree Accessibility Measures

The most basic measure used in space syntax is the step length and here we compute step lengths in their inverse form for the primal and dual problems. We will define these measures again as $\Delta_i^X = 1/\sum_k D_{ik}^X$ and $\Delta_j^Y = 1/\sum_\ell D_{j\ell}^Y$ and we show their form in Figures 7(a) and (b). This is much more intuitively satisfying as a representation of nearness between node centroids in the primal and streets in the dual, and a visual comparison of nodes with streets seems to confirm that these patterns of accessibility are close to one another. However what these measures reveal is that first there are profound edge effects which have probably been

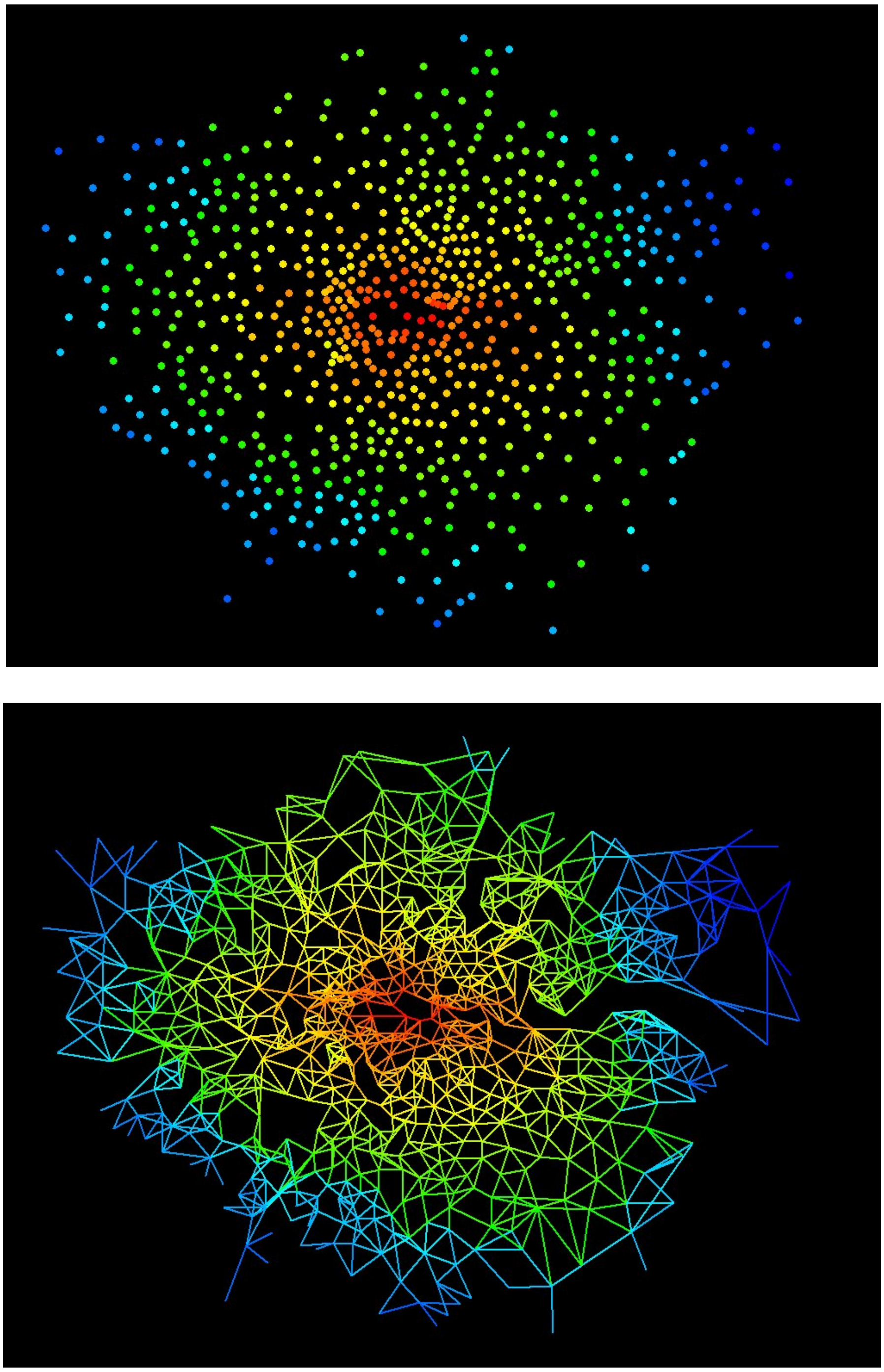


Figure 7: Nodes (a) and Street Accessibilities (b) Based on the Inverse Step Length Accessibility Measures

exacerbated by the way we have built the network to its edges – that is, the edge nodes are simply in the network so that they can meet the requirement of each non-edge node having at least 5 links to other nodes. Second, the fact that we deal with what is essentially a circular system means that the most accessible points are towards the centre of the system. This is a generic problem in all spatial analysis, and it relates to the basic issue of closing the system at some point to the outside world. Third there is the issue of local versus global connectivity in such a network and it clear that the more links that are taken into account, the more the structure at its most local scale is compromised. If we compare Figures 6 and 7, then we can see how local structure evolves to global structure as the measure of accessibility is based on wider and wider link effects. In fact for the step lengths, the number of matrix powers that is needed to span the entire system is some 29 for nodal connectivity and 27 for street connectivity. The computation of these step lengths using the matrix power method as implied by equations (4) to (7) is quite time-consuming for 699 x 699 and 1972 x 1972 matrices (all night on my PC Vaio VPCZ21M9E) and when we reach the point where it takes at least 29 or 27 steps to reach any and every node or street, this is still an arbitrary cut off. In fact it is worth showing how fast this computation is noting that at any point up to $n = 29$, we could take the step length matrix as a basis for the computation of accessibility measures. Of course when we reach the point where the cells of the step length matrix are positive, then we would need to work with the total numbers of path lengths, weighted or otherwise, as in equation (8). We show these trajectories in Figure 8.

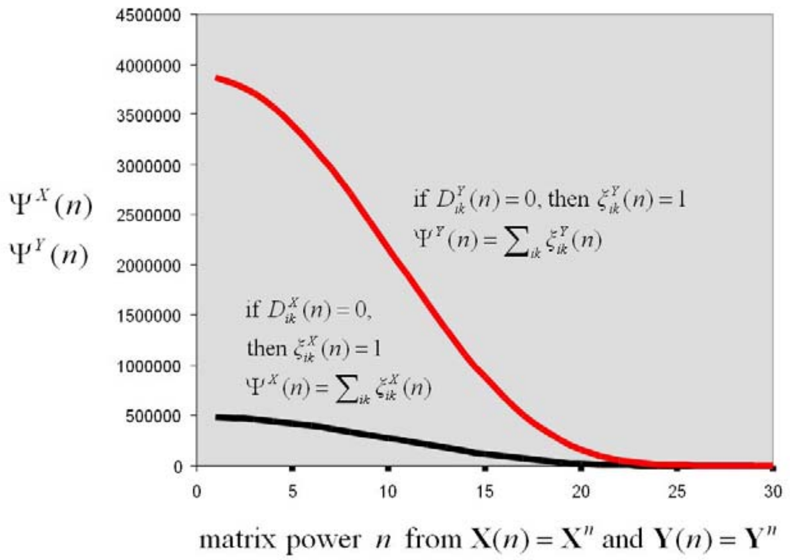


Figure 8: Convergence to the Step Length Limit for the Primal and Dual Problems

To see how these make a much bigger difference, we turn to our last exploration of accessibility in the primal and dual problems where we formulate the problem in probability terms. The steady state equations which we defined earlier as $\hat{\mathbf{p}} = \hat{\mathbf{p}}\hat{\mathbf{P}}$ and $\hat{\mathbf{q}} = \hat{\mathbf{q}}\hat{\mathbf{Q}}$ define processes where a walker starting from any position in the system – in the primal from any node and in the dual from any street, moves from node to node or from street to street with the probabilities of moving from one to another gradually reflecting the overall structure of nodes or streets with the initial probabilities washing out, diffusing if you like. In a system with very little structure which to an extent is our example – and this means we need a much better and fuller test of these ideas – then the probabilities of each node or street in the steady state are likely to be fairly similar. In short these Markov processes wash away the original probabilities and what remains is the true structure. We illustrated in earlier examples quite a high correlation between the steady states and the local structure but these were very simple graphs with exaggerated structure. Where one has large swathes of metropolitan area with similar structure in terms of the street network, then it is likely that the steady state is somewhat less distinct then the step method of accessibility just illustrated.

We show the node and street accessibility patterns based on $\hat{\mathbf{p}}$ and $\hat{\mathbf{q}}$ in Figure 9(a) and (b). Because the nodal structure is quite flat, we have scaled the values and then ranked them as we show in Figure 9(c) but this does little to sharpen the structure. In fact in Figure 9(b), the pattern is completely flat for the accessibility of streets with the edge nodes soaking up the probabilities in an obscure manner. It is worth comparing these patterns formally and in Figure 10 we have re-plotted the nodal patterns as thematic maps where each centroid is associated with each zone. This makes the patterns much easier to grasp intuitively. It is quite clear that the in-degrees are identical to the probabilistic steady state vectors with a correlation of 1 (Figure 10(a) compared to 10(c)). The correlation between the in-degrees and the inverse step length (10(a) cf 10(b)) is modest at 0.48 while that between in-degrees and the ranked probabilistic steady state vectors is 0.86. If we examine the dual street patterns, we have similar correlations but we have not ranked the street lines. The strongest correlation is low at 0.34 between the in-degrees and the inverse step lengths while that between the probabilities as inverse step lengths and the probabilities and in-degrees are both negative but less than 0.25. This as we have argued above is due to the nature of the diffusion of probabilities on the particular localised street graph that we have used.

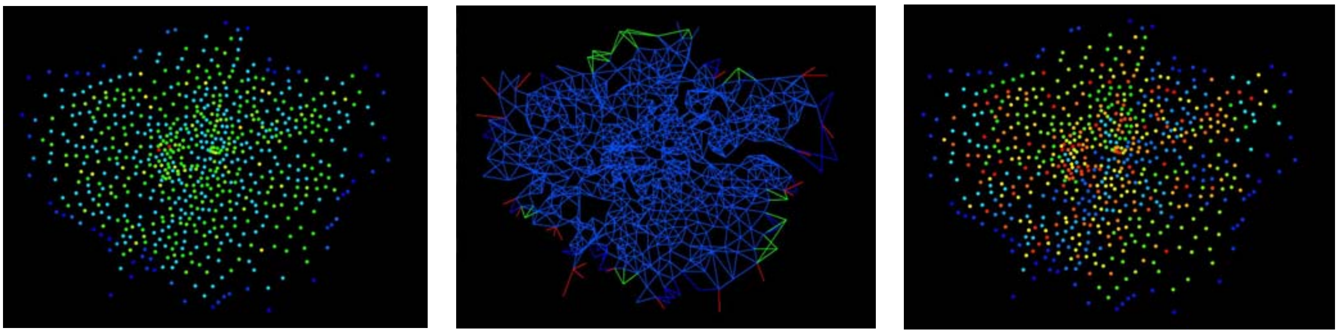


Figure 9: Steady State Nodal and Street Structure a) absolute values of $\hat{\mathbf{p}}$
b) absolute values of $\hat{\mathbf{q}}$ *c) ranking of* $\hat{\mathbf{p}}$

Our last issue with respect to the measures developed so far is to provide a partial test of how good the accessibility values are where we are able to match them against activities/trips associated with nodes and streets. The easiest test is to see how close the nodal accessibility values are to the observed activity totals associated with the set of centroids. These observed activity totals are origin employment and resident working populations which are formed from

$$\left.\begin{aligned} E_i^{obs} &= \sum_k T_{ik}^{obs} \\ P_k^{obs} &= \sum_i T_{ik}^{obs} \end{aligned}\right\} \quad . \qquad (23)$$

There is an immediate issue in terms of making comparisons at this level for it is not clear if the various access measures are more related to explaining employment or population. In fact there is a negative correlation of -0.217 between employment and population which is quite consistent with the spatial structure of largely monocentric cities with low population densities and high employment densities at and around their centres. There is another issue. The zonal structure is organised so that populations in each zones are as close to one another as possible. Although this is not strictly enforced as in the US where redistricting of electoral districts takes place after each election, there is momentum to make sure that there are no big differences between the electoral population in each ward. This means that a measure such as the inverse step length $\boldsymbol{\Delta}^X$ which from Figure 7(a) which increases as one gets closer to the centre would not explain the spatial structure of population which is more uneven.

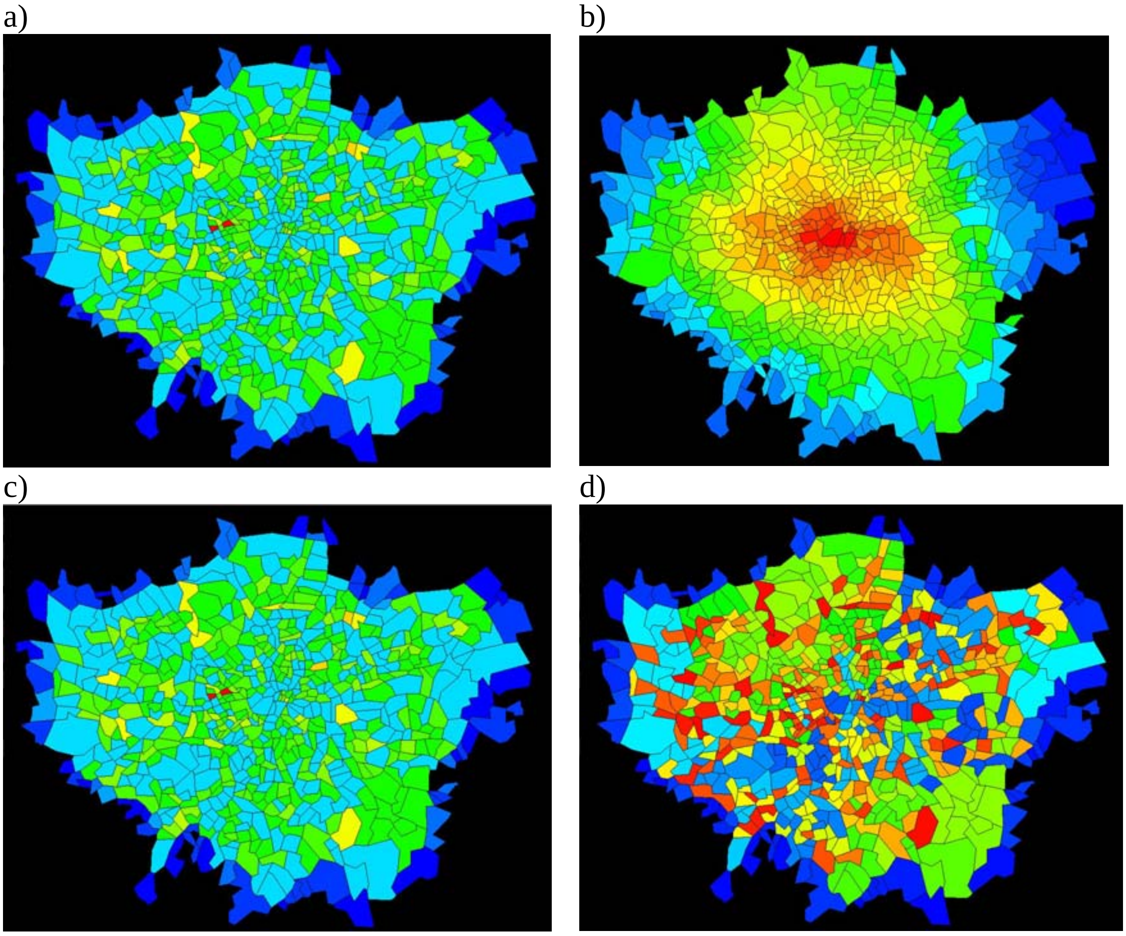


Figure 10: Comparisons of Nodal Accessibility Vectors
a) In-degrees b) Inverse Step Lengths c) Steady State d) Ranked Steady State

So what we do here to normalise these spatial equalities is to compare the accessibility vectors to the employment and population densities which we compute as $E_i^{obs} / Area_i$ and $P_k^{obs} / Area_i$ and use these to make comparisons with the accessibility vectors. The correlation between these densities is still low with virtually no correlation at 0.055. Nevertheless the predictions are better than expected. In fact we compare only the 633 zones in terms of the accessibilities and activity density vectors leaving out the 66 external edge of area zones. The two correlations between inverse step length and employment and population are both positive with population higher at 0.555 than employment at 0.371 and if we then compare them with the in-degrees $\mathbf{x}$, these correlations are much lower with no significance. The probabilistic measure from the steady state accessibilities $\hat{\mathbf{p}}$ is equivalent to the in-degrees. In fact we consider the correlations with the inverse step lengths to be significant and when we examine plots of these values, it is clear there are positive relationships with employment having a very characteristic scatter which is almost super-exponential. We show these plots in Figures 11(a) and (b). To an extent, it is a little surprising that our measures correlate so well with densities for the underlying accessibility measures based on crude step lengths whose basic data range from 1 to 29 in value are relatively unsophisticated indices.

a) b)

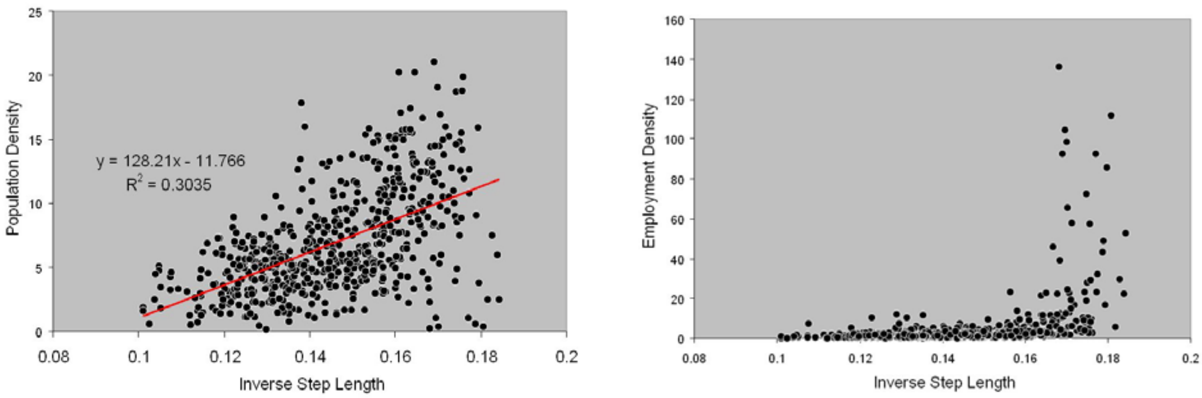


Figure 11: Empirical Comparisons of the Inverse Step Length with a) Population Density and b) Employment Density

We can now explore the question of observed movements on the space syntax street links. In fact, we know *a priori* that we do not have the actual trip movements on each link for all we have are generic interactions between origins and destinations that in fact have to be mapped or assigned to the network before we can produce actual trips on the network. Thus it is likely that any analysis of origin-destination trip movements on specific network links is likely to be flawed. For each street j in the system which is associated with a link between intersections ik, we can compare the accessibility on that link with respect to the flow between an origin and destination from the observed matrix of flows T_{ik}^{obs}, notwithstanding the fact that these are quite different from the actual flows. The problem as we have pointed out earlier is that the observed values we have are not those that are actually observed on any link ik for these values combine many trips between origins and destinations which are assigned to the link in question. The data too is only a 10% sample from the 2001 Population Census. Moreover the actual network structure that we have developed is a nearest neighbour network and it does not include long links such as motorways and other major roads with restricted access. If our threshold on links were to be relaxed and the notion of streets with more than two intersections with other streets to be invoked, we might improve the comparison but this requires testing and further development on a much richer and more detailed network. In terms of the correlations, the in-degree and inverse step lengths have barely any correlation with observed trips at -0.073 and -0.097 (noting that the in-degree and inverse step lengths correlate at 0.342). The correlation between observed trips and the probabilistic access measure is actually very slightly negative at -0.064 but essentially there is no correlation. Despite these results being somewhat disappointing, they are entirely explicable in terms of the data used and the fact that these data does not contain anything other than nearest neighbour links to explain urban structure. In the next and last section of the paper, we will explore a way forward.

An Integrated Approach

The key to an integration of space syntax and spatial interaction has already been defined through the two operations on the basic matrix $\mathbf{A}$ which give rise to the primal and dual interaction matrices $\mathbf{U} = \mathbf{A}\mathbf{A}^T$ and $\mathbf{V} = \mathbf{A}^T\mathbf{A}$. The distances at different step lengths although related from $\mathbf{U}^n$ and $\mathbf{V}^n$ by $\mathbf{U}^n\mathbf{A}^T = \mathbf{A}\mathbf{Y}^n$, have to be normalised for interpretation. However if we work with the dual normalisation of the basic matrix as row stochastic probability matrices $\mathbf{G} = [G_{ij}]$ and $\mathbf{C} = [C_{jk}]$ from equations (19) and (20) which we restate below as

$$G_{ij} = \frac{A_{ij}}{\sum_j A_{ij}} \quad , \quad \sum_j G_{ij} = 1 \quad , \qquad \text{and} \qquad [(19)]$$

$$C_{jk} = \frac{A_{jk}^T}{\sum_k A_{jk}^T} \quad , \quad \sum_k C_{jk} = 1 \qquad , \qquad [(20)]$$

then successive powers of the probability matrices $\hat{\mathbf{P}} = \mathbf{GC}$ and $\hat{\mathbf{Q}} = \mathbf{CG}$ give very clear steady state relations $\hat{\mathbf{p}}\mathbf{G} = \hat{\mathbf{q}}$ and $\hat{\mathbf{p}} = \hat{\mathbf{q}}\mathbf{C}$. We have, however, demonstrated that these steady state relations wash away the structure that we need to preserve as a key determinant of the relevant accessibilities of nodes/centroids and streets, thus we begin with the matrices $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$.

Thus a more basic approach is to assume that the probability matrix $\hat{\mathbf{P}}$ is the determinant of the singly constrained trip equation which we stated earlier in equation (13) and now elaborate as

$$\left.\begin{aligned} T_{ik} &= E_i p_{k|i} = E_i \hat{P}_{ik} \\ E_i &= \sum_k T_{ik} = E_i \sum_k \hat{P}_{ik} \\ \rho_k &= \sum_i T_{ik} = \sum_i E_i \hat{P}_{ik} \end{aligned}\right\} \quad \text{where} \quad \sum_k \hat{P}_{ik} = 1 \qquad . \qquad (24)$$

Without elaborating the density version of the model as above, we will substitute employment density $E_i / Area_i$ for the employment count and test both counts and densities in the following application. We know that the $\hat{\mathbf{P}}$ matrix in the examples so far in this paper is very sparse as it is a nearest neighbour network but if we assume it is sufficiently rich to detect urban structure, then the spatial interaction model follows directly from equations (24). Note that the vectors $\mathbf{e} = [E_i]$ and $\boldsymbol{\rho} = [\rho_k]$ are not the steady state vectors but origin and estimation vectors which we can interpret in spatial interaction terms as employment and population. In matrix terms we write the model in equation (24) as

$$\boldsymbol{\rho} = \mathbf{e}\hat{\mathbf{P}} = \mathbf{eGC} \qquad (25)$$

which is the primal spatial interaction and then by applying the matrix $\mathbf{G}$ to this equation, we generate the dual space syntax model as

$$\boldsymbol{\rho}\mathbf{G} = \mathbf{e}\hat{\mathbf{P}}\mathbf{G} = \mathbf{eGCG} = \mathbf{eGQ} \qquad . \qquad (26)$$

$$\mathbf{r} = \mathbf{sQ} = \boldsymbol{\rho}\mathbf{G} = \mathbf{eGQ} \qquad , \qquad (27)$$

where it is now clear that $\mathbf{r}$ and $\mathbf{s}$ are the equivalent to population and employment (counts or densities) respectively but now spread to the street network; that is, $\mathbf{r}$ and $\mathbf{s}$ are the population and employment equivalents that simply relate these to the streets. What this means is that population and employment are spread from locations to their connected street lines. The predictions of $\boldsymbol{\rho}$ and $\mathbf{r}$ are thus entirely consistent with one another and can be derived from one another as equation (27) reveals.

We have tested equations (25) to (27) on our Greater London network and essentially what we do is take the employment for each location and work out the population using the matrix $\hat{\mathbf{P}}$ as in equation (25). We do this for both the count and density employments and the data that we use is shown in Figure 12.

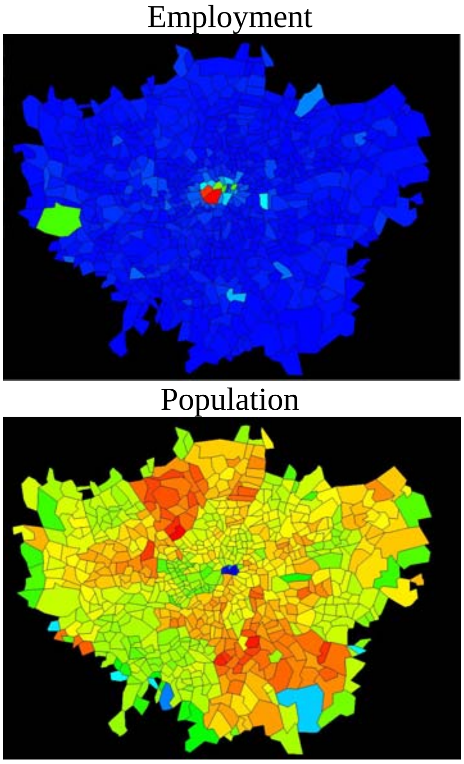


Figure 12: The Distribution of Aggregated Trips at Origins (Employment) and at Destinations (Population) as Counts and Densities

The relative concentrations in Figure 12 are consistent with the fact that the population (and its density) are much more spread out than employment counts and densities. In essence, what the model does is translate in primal form, the employment counts and densities in the upper row of Figure 12 to their population equivalents in the lower row using the matrix $\hat{\mathbf{P}}$. We show these predictions in Figure 13 where we plot counts and densities in the upper row and show their form in the lower row through ranking which reduces the spread of these thematic maps.

It is very clear that the translation from employment counts and employment densities uses a probability matrix which has so little structure within it – from the raw planar graph – that it hardly translates employment into population, the results for both counts and densities being very close to the original distributions of employment. If you compare the employment maps in Figure 12 with the population in the upper row of Figure 13, the correlation between the two is very high. The correlations between predicted and observed population counts is negative at -0.150 while for densities, it is positive but low at 0.119. To an extent, this reflects the major conclusion of this work: that in many space syntax analyses, because the planar graph used is one based largely but not exclusively on nearest neighbours, there is not enough structure in this matrix to ensure that we get good predictions of locational activities which in turn are derived from trip movements. We will come back to this point as it is perhaps the most important finding from this analysis in that it reflects the notion that we need to think about space syntax in terms of other approaches and only then can we assess how appropriate the approach is.

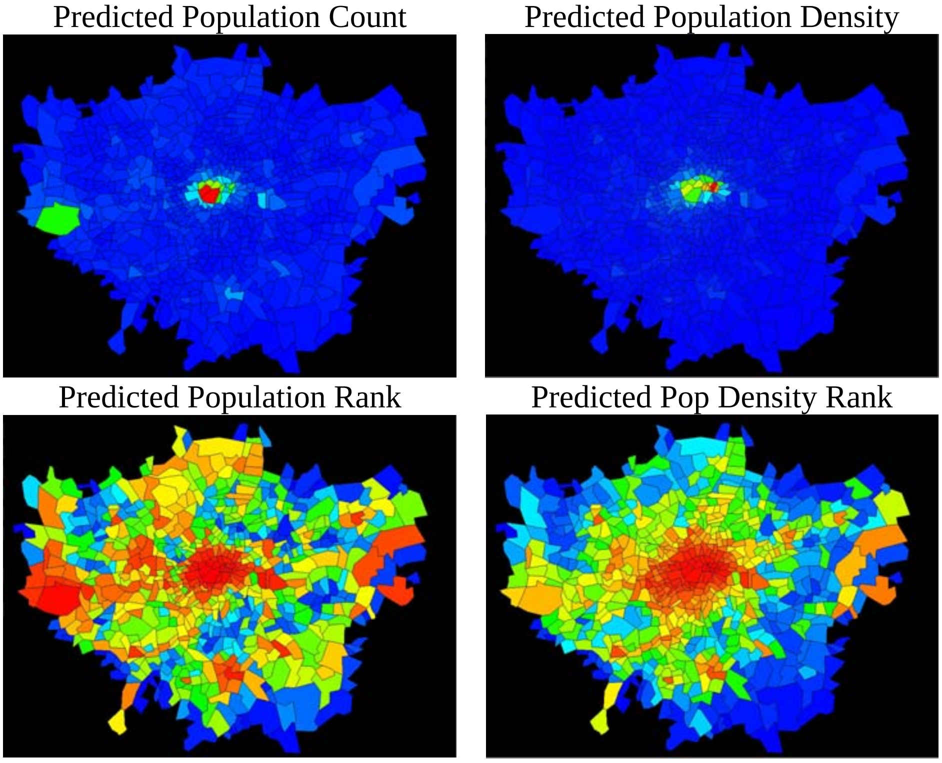


Figure 13: Predicted Population Counts and Densities from the Primal Interaction Model

In fact all is not lost even from this application, for when we rank the predicted population counts and densities, we do see some structure. Compare the observed population density with the ranked predicted population density – the bottom right hand map in Figure 12 with its equivalent in Figure 13 – and we see a much stronger correlation which shows that there is some structure in the matrix $\hat{\mathbf{P}}$. We can also get at this by comparing a logarithmic transformation of the predicted and observed densities as in Figure 14 which reveals a stronger significant correlation at 0.507. Doubtless, if we were to produce a more structured basic probability matrix – perhaps $\hat{\mathbf{P}}^2$, $\hat{\mathbf{P}}^3$, etc. –then it is possible we would get better results even with this simple and somewhat arbitrary example.

The last thing we will do is transform the primal spatial interaction model into its dual space syntax equivalent. Equations (26) and (27) illustrate that it is a simple matter to convert employment and population activity at centroids or nodes into activity which is spread along the links of the system which are streets. This population in fact is $\mathbf{r} = \boldsymbol{\rho}\mathbf{G}$ while the employment is $\mathbf{s} = \mathbf{eG}$. We could, of course, had we estimates of these activities associated with streets, start with employment spread along streets using $\mathbf{Q}$ to predict $\mathbf{r}$ from $\mathbf{s}$ but there is no tradition of working in this manner. It is not out of the question however to begin to collect activity along streets and pursue the analysis in this direction. If our streets are longer segments with more than two intersections, this makes the analysis more convoluted but it is still possible to imagine there are insights into urban structure to be achieved in this way. To conclude, we show the street flows of population activity $\mathbf{r}$ and employment activity $\mathbf{s}$ for both counts and densities in Figure 15. It is very clear that employment dominates these spreads and because the basic probability matrices are so sparse and simply connected, the spread to population also mirrors employment. It is possible if we rank these links rather than use absolute values, that more structure could be extracted from these patterns. But this is just one of many explorations that we could continue to make . However we consider that we have now pointed the direction and that a number of lines for future research have been established. We turn to these by way of conclusion.

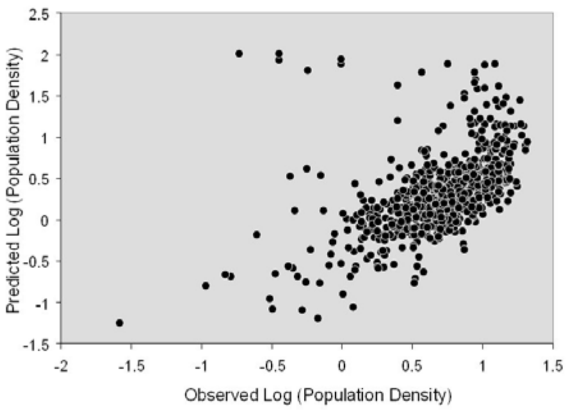


Figure 14: Logarithmic Predicted and Observed Population Densities

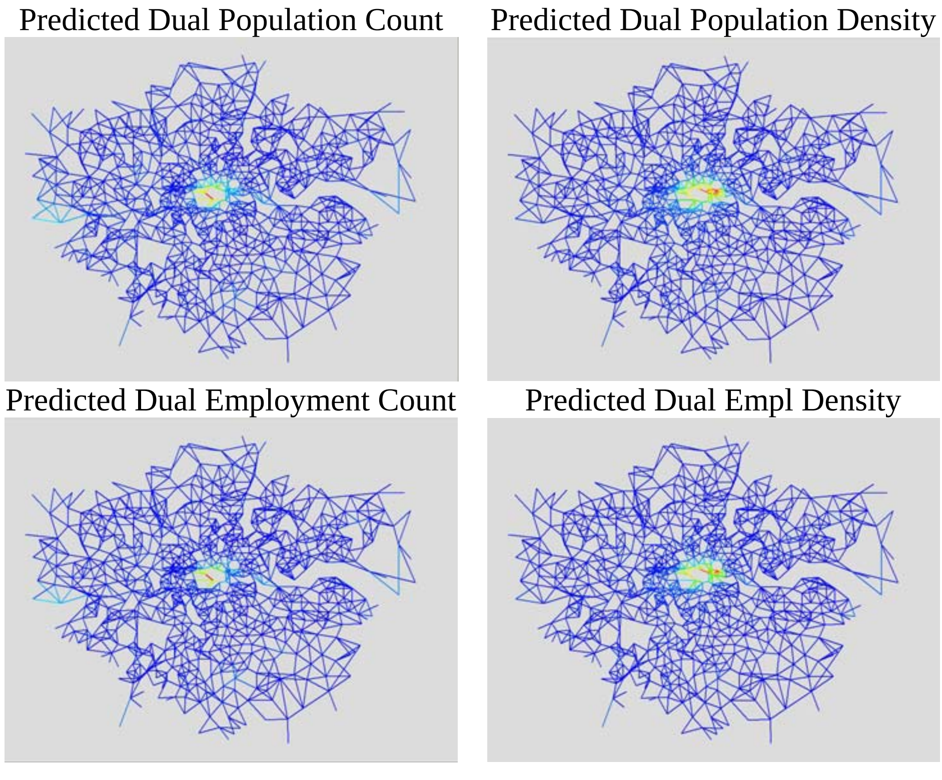


Figure 15: Observed Employment and Predicted Population Counts and Densities for the Space Syntax Dual Formulation

Conclusions: Next Steps

The key issue in predicting urban movements in spatial interaction models involves the independent variables which represent the trade-off between measures of the size of locations and the cost or distance from a location where movement is generated and a location to which it is attracted. All of this information is represented in the probability matrix $\hat{\mathbf{P}}$ which we have articulated in primal form as the interaction between the set of centroids/locations and the streets or routes to which they are linked, that is using the matrix $\mathbf{A} = [A_{ij}]$. The primal matrix is dimensioned to represent all possible movements between the nodes which is of the order N^2 and this in turn is based on the number of street links or segments L which we have suggested is very much smaller than the total number of movements. How good this structure of streets is in representing all the nuances and biases in urban structure depends to a large extent on this number. In the application to Greater London including external zones, there are $N = 699$ nodes with a possible number of trip movements 488,601 $(= N^2)$ whereas the nearest neighbour street network has some 3,944 $(= L)$ which means that the number of possible links on which trips might be observed is only 0.008, not quite 1%. Were we to increase the number of possible links to N^2, then we would need to consider links between all these possible streets. In fact, it is most unlikely that all possible trips would use all possible links for different trips are assigned to the local street segments in making a shortest route between origins and destinations.

The way we have represented the problem in terms of a primal matrix with dimensions N^2 and the dual matrix with dimensions L^2, involves non-trivial matrix computations in terms of size. Were we to have as many segments as possible origin-destination interactions, our dual matrices would be of the order $N^4 = N^2 N^2$ which in the London example would be $488{,}601 \times 488{,}601$ giving matrices with some 23,873,093,7201 (23 billion) cells. Such matrices, frankly, are simply beyond our capability to work with. However, it is most unlikely that every street segment relates to every other which is what this would imply, and thus this number of cells is a theoretical upper limit. But what it does show is that it is absolutely essential to get the structure of the basic graph correct and it is quite clear that in this application, we have far too sparse primal and dual matrices. This suggests that we need to pay particular attention in space syntax to the nature of the street matrix; and it also suggests that if we are to link this to spatial interaction, we need to define the $\mathbf{A}$ matrix in much richer terms, taking account of size as well as connectivity.

Throughout this paper, we have been at pains to state that the configuration of the basic planar graph from which the dual and primal interaction matrices are defined is critical to appropriate applications of both of these approaches and particularly their integration. What we now need are better examples with richer structure and then we will be able to assess the extent to which our measures of accessibility and the integrated model posed in the last section can be developed further. We also need to explore the extent to which the spatial system which is represented at the zonal and street scales can be reconciled and this probably means that we need to consider how trip movements are assigned to street segments. This might in fact be a good criterion for defining the connectivity matrix in the first instance but it also requires considerable further research to bring this kind of analysis to fruition. Last but not least, we need to say something about whether or not we have made progress with our integration of space syntax and spatial interaction here. What is clear is that we have clarified considerably how we might develop any such integration but our applications have been disappointing in that our example is not rich enough to show good results involving prediction. This is very much reflected in the data we have used, particularly the street network but it has enabled us to say something very significant about how we define the networks used in space syntax. To progress these to the point where they are useful for spatial interaction models, we must devise much clearer rules for the definition of the basic network, its topology, and its density on which the various accessibilities we have defined are to be measured. Only then will we be able to progress space syntax to the point where it is consistent with the use of spatial interaction modelling in prediction.

References

Bafna, S. (2003) Space Syntax: A Brief Introduction to Its Logic and Analytical Techniques, **Environment and Behavior**, **35**, 1, 17-29. DOI: 10.1177/0013916502238863.

Barthélemy, M. (2011) Spatial Networks, **Physics Reports**, **499**, 1-101, also at arXiv: 1010.0302 [cond-mat.stat-mech]; DOI: 10.1016/j.physrep.2010.11.002.

Batty, M. (2013) **The New Science of Cities**, The MIT Press, Cambridge, MA.

Batty, M. (2016) Evolving a Plan: Design and Planning with Complexity, in J. Portugali and E. Stolk (Editors) **Complexity, Cognition, Urban Planning and Design**, Springer Proceedings in Complexity, Springer International, Zurich, Switzerland.

Borgatti, S. P. and Everett, M. G. (1997) Network Analysis of 2-Mode Data, **Social Networks**, **19**, 243-269.

Hansen, W. G. (1959) How Accessibility Shapes Land Use, **Journal of the American Institute of Planners**, **25**, 73-76.

Hillier, B., and Hanson, J. (1985) **The Social Logic of Space**, Cambridge University Press, Cambridge, UK.

Hillier, B. (1996) **Space is the Machine**, Cambridge University Press, Cambridge, UK.

Hillier, B., Penn, A., Hanson, J., Grajewski, T., and Xu, J.(1993) Natural Movement: Or, Configuration and Attraction in Urban Pedestrian Movement, **Environment and Planning B**, **20**, 29–66.

Marshall, S. (2015) Line Structure Interpretation of Networks: Relationships, Matrices and Properties, Working Paper 204, CASA, University College London WC1E 7HB, UK

Wilson, A. G. (1970) **Entropy in Urban and Regional Modelling**, Pion Press, London.

Acknowledgements

The author thanks the Chinese University of Hong Kong for funding his appointment as VC Distinguished Visiting Professor during which this paper was written (January 2017). This research has also been supported by the EPSRC UK Regions Digital Research Facility (UK RDRF) Grant Number EP/M023583/1.