Network effects and psychological effects: a theory of urban movement

Bill Hillier and Shinichi Iida
University College London, UK
b.hillier@ucl.ac.uk, s.iida@ucl.ac.uk

Abstract

Correlations are regularly found in urban movement between configurational measures of the axial graph and observed movement patterns. This suggests that topological and geometric complexity are critically involved in how people navigate urban grids. This has caused difficulties with orthodox urban modelling, since there it has always been assumed that insofar as spatial factors play a role in navigation, it will be on the basis of metric distance. In spite of much experimental evidence from cognitive science that geometric and topological factors are involved in navigation, and that metric distance is unlikely to be the best criterion for navigational choices, the matter has not been convincingly resolved since no method has existed for extracting cognitive information from aggregate flows. Within the space syntax literature it has also remained unclear how far the correlations that are found with syntactic variables at the level of aggregate flows are due to cognitive factors operating at the level of individual movers, or are simply mathematically probable network effects, that is emergent statistical effects from the structure of line networks, independent of the psychology of navigational choices. Here we show how both problems can be resolved, by showing three things: first, how cognitive inferences can be made from aggregate urban flow data and distinguished from network effects; second by showing unequivocally that urban movement, both vehicular and pedestrian, are shaped far more by the geometrical and topological properties of the grid than by its metric properties; and third by demonstrating that the influence of these factors on movement is a cognitive, not network, effect.

1. Introduction and summary

A fundamental proposition in space syntax is that, with the kinds of exceptions noted in Hillier et al (1987), Hillier et al (1993), Chang & Penn (1998) and Penn et al (1998), the configuration of the urban street network is in itself a major determinant of movement flows. This seems to imply that topological and geometric complexity are critically involved in how people navigate urban grids, but because the reported results, however, are about aggregate human behaviour, it has always been unclear how far they depended on individual spatial decisions, and how far they are simply mathematically probable network effects, that is emergent statistical effects from the structure of line networks, relatively independent of the psychology of navigational choices.

The apparent involvement of grid complexity in navigation has also brought space syntax into conflict with orthodox urban modelling, where it has always been assumed that insofar as spatial factors play a role in navigation, it will be on the basis of metric distance. However, in recent years, research results have accumulated in cognitive science which suggest that the metric distance assumption is unrealistic, not perhaps because we do not
seek to minimise travel distance, but because our notions of distance are compromised by the visual, geometrical and topological properties of networks. For example, estimates of distance have been shown to be affected by the division of routes into discrete visual chunks (Golledge 1992, Montello 1997, Kim 2001), by a tendency to correct bends to straight lines and turns to right angles (Allen 1981), and even by the direction in which the estimate is made (Sadalla, 1980, Montello 1992, Golledge 1995). As a consequence, much current cognitive work on spatial complexity explores how far route choices reflect the frequency (Duckham 2003) or degree (Conroy-Dalton R 2001, 2003, Hochmair 2002) of directional change, rather than metric distance, reflecting current choices in space syntax (as for example in WebMap which allows angular as well a topological relations between the nodes of the axial graph).

An obstacle to a more definitive resolution within the urban research community of how distance concepts shape human movement - or even whether or not a general definition exists - is that no method exists to extract cognitive information from the aggregate flows in street networks, and distinguish this from emergent statistical effects of the network itself. In this paper, we seek to resolve these questions through a two stage argument. First we develop a theory of why network effects on movement flows are to be expected in spatial configurations in general, and why the syntactic measures of integration (more commonly called closeness outside space syntax) and choice (commonly called betweenness) can be expected to capture them.

We then ask what remains for the psychology of navigation and suggest the answer can be found in the concepts of distance that that must underlie the use of measures like integration and choice. By using different concepts of distance in configurational analysis of urban networks, and correlating the results with real movement flows, we show how cognitive inferences can be made from aggregate movement data, and distinguished from network effects. By using this method in a study of movement in four areas of London, we also show unambiguously that movement in cities reflects the geometrical and topological structure of the network configuration far more than metric distance. It is therefore not the case, as has been argued as self-evident, that weighting axial graphs for metric distance will improve correlations with observed movement (Steadman 2004). On the contrary, such a weighting virtually eliminates the ability of the integration measure to predict, and severely reduces it in the case of the choice measure.

2. Network effects: theoretical motivation

Why and how, then, should we expect street networks to shape movement in cities? First, we must be clear about network effects - that is emergent statistical effects on aggregate movement from the structure of the network itself - and why they are to be expected in movement in urban systems. First, we remind the reader of a simply motivating example (Hillier 1999) that makes network effects intuitively obvious.

Figure 277 shows a notional grid with a main street, a cross street, some side streets and a back street. Suppose all streets are equally loaded with dwellings, and that people move over time from all dwellings to all others using some notion of shortest (least distance) or simplest (fewest turns) routes. It is clear that more movement will pass through the horizontal main street than other streets, and that more will pass the central parts of the main street than the more peripheral parts. This effect follows from the structure of the network, since no one need plan to pass through the spaces, and would hold under either
assumption about distance. It is also clear that the main street considered as a whole is more accessible than other spaces, on either definition of distance. It will then be more advantageous to locate a shop on the main street, since it will be both easier to get to and also where people are likely to be when moving between locations. Although locating a shop is an individual decision, it is clear that the decision will be shaped first and foremost by the properties of the network. This simple example shows in a commonsense way why we should expect network effects on movement.

Now consider a more complex theoretical example. On the top left of Figure 278 is a notional arrangement of blocks with something like the degree of linear continuity between spaces that we expect in urban space.

Visual integration analysis through DepthMap shows an emergent warm colour pattern which looks a bit like a main street, with side streets and back streets, although of a rather irregular kind. On the right we retain exactly the same blocks but marginally move some to ‘just about’ block lines of sight, so in effect, we change nothing but the linear relations between some of the spaces. The visual integration analysis, shown on the same scale, shows both a substantial loss of visual integration but also the pattern is totally transformed, with little in the way of continuous structure. Spatial network seems to have lost structure as well as integration.

This is of course the well-known syntactic property of intelligibility. The $r^2$ between the visual connectivity and visual integration of point is 0.714 in the left case and 0.267 on the right. But this mathematical change in the network structure has consequences even for theoretical movement. In bottom of Figure 278, we use the ‘agent’ facility in Turner’s DepthMap software to show the traces of 10000 sighted agents with 170 degrees of vision, who select a point within their field of view randomly, move towards it three pixels and the repeat the process. Even with the sight and distance parameters set close to randomness (in that vision is diffuse rather than focused and distance between destination selections are short) the results are strikingly different. In the left case the highest density of traces follows the space structure to a remarkable degree, while in the right case, it spirals all over the system, reflecting the local scaling of spaces rather than overall configuration. Since the ‘cognitive’ structure and behaviour of agents is identical in the two cases, the differences between the trace patterns are clearly network effects.
3. Network effects in real urban systems

But what of real human subjects in real urban systems? Are there reasons for expecting network effects at this level? It is useful here again to distinguish between the structure of the graph, that is, the pattern of nodes and links, and how distance between nodes is to be calculated. There are (as noted in (Hillier 2002) in principle strong mathematical reason to expect network effects from the structure of any graph on movement, in that random one step movement in a non-bipartite graph will lead to the number of visits per node going to a limit of the degree of the node as the number of iterations goes to infinity. (Norris 1997, Batty & Tinkler 1979). If movement is random and one step, then, the number of visits to each node will be wholly determined by the structure of the graph.

But of course human movement is neither random nor one step. For the most part it is both planned and n-step. Are there also mathematical reasons why flows arising from this kind of movement will be shaped by the network configuration? First let us consider the nature of human movement. It has two aspects: the selection of a destination from an origin; and the selection of the intervening spaces that must be passed through to go from one to the other. The former is about to-movement, the latter through-movement.

First, consider to-movement. From any origin, say, someone’s house, we must expect that over time a range of trips will be made to various destinations, and these will be a matter of individual decision. But, over time, the choice of destinations from any origin would be expected to show some degree of statistical preference for closer rather than more remote destinations - say, by going more often to the local shop than to visit an aunt in Willesden. It does not have to be that way, but in most cases it probably will be. This is no more than an instance of what geographers have always called distance decay. But if there is any degree of distance decay in the choice of destinations, then it has the formal consequence that locations which are closer to all others in the network, will feature as destinations more often than those that are more remote - that is, more
accessible locations will be theoretically more attractive as destinations than less accessible ones simply as a result of their configurational position in the complex as a whole. The bias towards more accessible locations for to-movement is then a network effect, due to the configurational structure of the network, even though it is driven by the accumulation of individual decisions.

This of course is to say no more than that central locations in a system are more accessible than others. But the argument becomes more interesting if we consider variable radius integration, that is the accessibility of nodes to its neighbouring network up to a certain graph or metric distance away. A node which is more integrated than others in its region at a given radius will also become more theoretically attractive as a destination to the degree that movement at that graph scale is preferred in the system. In other words, the network properties measured by variable radius integration can be theoretically expected to attract more movement to some destinations that others purely as an effect of the structure of the network, and this is of course what we find in urban reality.

But the effect of the network does not end there, since there will also be network effects on through-movement, more obvious than those for to-movement. The sequences of nodes that are available between origins and destination will often vary with different definitions of distance (as we will see below) but whichever definition of distance we choose the available sequences are defined by the structure of the graph itself, so again we are dealing with network effects. However we define distance, in effect, the choice of routes is defined by the network.

So network effects must exist for both to and through movement. There will also be interaction between them. First, every trip is made up of a pair of origin-destination, or to-movement nodes, and a variable number of through-movement nodes. But with increasing length of trip the to-movement pair will remain constant at two nodes while the through movement node count will increase. It follows that the longer the trip, the higher will be the ratio of through movement spaces to the origin destination pair, which of course always remain constant. We may expect then that the greater the graph length of the trip, the more it will reflect the choice, or betweenness, structure of the graph, rather than the integration, or closeness, structure.

Second, any to-movement bias towards integrated locations will also have an effect on through-movement, since routes leading to those locations will be more likely to be used than those leading to less integrated locations. It would be a simple matter to reflect this by weighting the choice measure for the integration value of destinations (see below). This combined measure should then theoretically measure both aspects of simplest path n-step movement in a system, with some adjustment for the mean length of trips. So in the ‘state of nature’ the graph should already have a tendency towards a certain pattern of movement reflecting the spatial configuration of the graph, and the configurational properties which produce this are exactly reflected in the syntactic measures of variable radius integration and choice, and by the relations between them.

This then is the theory of urban movement in a network considered as a graph. But how people actually move will clearly be affected by how distance is conceptualised. In what follows, we show how cognitive information on how distance concepts can be extracted from information on real flows in urban networks.
4. Varying distance concepts in line networks

The technique we propose to extract cognitive information from real flows is to take urban street networks and subject them to different mathematical interpretations according to how distance is defined, then to explore how well the different interpretations correlate with real movement patterns. The basis of the different interpretation is the SEGMEN model developed by Shinchi Iida which is explained fully in (Iida & Hillier 2005), following earlier work by Turner, Dalton, Conroy-Dalton, Peponis and others (Turner 2001, Dalton 2001,2003, Dalton, Conroy-Dalton, Peponis 2003)

In the SEGMEN model, we start from the existing fewest line map and represent the street network as its graph of line segments between intersections. Three different weights are assigned to relations between adjacent segments: metric length, directional change, and degree of angular change, as in Figure 3, reflecting different conjectures in the literature as to how distance is conceptualised in human navigation.

Paths between all segments and all others can then be assessed in terms of least length, fewest turns, and least angle paths. Least length paths are the shortest metric distances, fewest turns paths the least number of direction changes, and least angle paths the smallest accumulated totals of angular change on paths, between all pairs of nodes. Note that the original lines of the fewest line map will emerge as sequences of zero-change weights for fewest turns and least angle change paths. In this sense the model allows also a test how far the focus on linearity of syntactic axial mapping is justified.

We apply the simplest and most generally used versions of the syntactic measures of integration and choice outside syntax to the weighted graphs, namely the common measure of ‘closeness’ and ‘betweenness’. Closeness, as defined by Sabidussi (25), is
\[ C_c(P_i) = \left( \sum_k d_{ik} \right)^{-1}, \]
where \( d_{ik} \) is the length of a geodesic (shortest path) between node \( P_i \) and \( P_k \). Betweenness, as defined by Freeman (26), is
\[ C_B(P_i) = \sum_j \sum_k g_{jk}(p_i) / g_{jk}(j < k), \]
where \( g_{jk}(p_i) \) is the number of geodesics between node \( p_j \) and \( p_k \) which contain node \( p_i \) and \( g_{jk} \) the number of all geodesics between \( p_j \) and \( p_k \).

This gives six different mathematical interpretation of a street system: closeness and betweenness measures applied to least length, least angle change and fewest turns weightings of relations between adjacent segments in the system. In addition, closeness measures can be assigned for every radius, defining radius in terms of the number of segments distant from each segment treated as a root. This permits experimentation with the scale at which measures operate, from the most local, to the global level.

5. Empirical studies

We then take four areas of London (Barnsbury, Clerkenwell, South Kensington and Knightsbridge) on which earlier studies (Penn et al 1998) had established dense vehicular and pedestrian movement flows at the segment level throughout the working day for a total of 356 observation ‘gates’. Two of the areas had originally been selected to pose problems for a purely configurational analysis, in that they had large movement attractors at their heart, one a complex of national museums adjacent to a tube station, and the other a leading department store. Both could be expected to distort correlations with purely network measures. The street network for each observation area was embedded in a contextual network of 3 - 3.5 km radius, and analysed using the segment network representation. Closeness measures were calculated for every third radius, that is up to 3, 6, 9... segments distant from a root segment, up to the maximum radius of the system.
Figure 279: Figure a is an unweighted line network with three lines. Figure b shows how the line network is disaggregated at intersections to form a segment network. Figure c shows its graph representation. Each line segment is represented as a node in the graph and links between nodes are intersections. The distance cost between two line segments is measured by taking a ‘shortest’ path from one to the other, so the cost of travel between S and a can be given as $w(\pi - \theta) + w(\varphi)$, while the cost between S and b can be $w(\theta) + w(\pi - \varphi)$. The following weight definitions are used to represent different notions of distance: Least length (metric): The distance cost of routes in measured as the sum of segment lengths, defining length as the metric distance along the lines between the mid-points of two adjacent segments. The distance of two adjacent line segments is thus calculated as half the sum of their lengths. Fewest turns (topological): Distance cost is measured as the number of changes of direction have to be taken on a route. In the example shown in Figure b and c, $w(\theta) = w(\pi - \theta) = w(\varphi) = w(\pi - \varphi) = 1$ (however, $w(0) = 0$). Least angle change (geometric): Distance cost is measured as the sum of angular changes that are made on a route, by assigning a weight to each intersection proportionate to the angle of incidence of two line segments at the intersection. The weight is defined so that the distance gain will be 1 when the turn is a right angle. In other words, $w(\theta) \propto \theta, (0 \leq \theta < \pi, w(0) = 0, w(\pi/2) = 1$
Figure 280: Figure 280 shows disaggregated line network models of Barnsbury area in London, each coloured up by a different measure. The upper row of figures show closeness (a: geometric, b: metric, c: topological) and the figures at the bottom betweenness (d: geometric, e: metric, f: topological). In the closeness maps (a-c), line segments in red are those with the least mean distance cost from that line segment to all others, and shift to blue as the degree of closeness declines. In the betweenness maps (d-f), line segments in red are those which lie on most ‘shortest’ paths between pairs of segments, declining towards blue for those which lie on least. The network consists of 10897 line segments and the area covered by this model is roughly 3km in radius. The white dotted lines indicate the observation area where vehicular and pedestrian flow data was collected. Of two measures, closeness shows a marked difference between different interpretations of distance, with the metric version yielding only a concentric picture with highest closeness in the centre of the map and a smooth decline towards the edges. Betweenness shows more stability across the weights, although a rather more concentric picture can be observed in the metric version.

Translating the numerical results of the analysis into images of the network, with segments coloured in bands of value for each measure, from red for least distance through to dark blue for most, we find first that the different interpretations give different pictures of the ‘structure’ of the network, some slight, others more substantial as in Figure 280.

Adjusted R-square values for the correlations between closeness and betweenness measures ranges between 0.505-0.807 for the least angle interpretation, between 0.599-0.853 for fewest turns, and between -0.0067-0.0764 for least length. The reason for the lack of correlation for the least length version is that least length for closeness from all nodes to all others must give a result in which the closest segment to all others is at the geometric centre, with a more or less smooth fall-off from centre to periphery, as can be seen in the concentric patterns shown in Figure 280b. The different sets of values for segments were then correlated with observed vehicular and pedestrian flows averaged for the whole working day between 8 a.m. and 6 p.m. For vehicular movement correlations were confined
Network effects and psychological effects: a theory of urban movement

Table 26: Adjusted R-square values for correlations between vehicular flows and least sum of length, least sum of angle change and least number of turns analysis applied to closeness and betweenness measures. Best correlations are marked *. Numbers in round brackets indicate best radius in segments for closeness measures.

<table>
<thead>
<tr>
<th>Area name</th>
<th>Gates</th>
<th>Measure</th>
<th>Least length</th>
<th>Least angle change</th>
<th>Fewest turns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barnsbury</td>
<td>116</td>
<td>closeness</td>
<td>0.131(60)</td>
<td>0.678(90)*</td>
<td>0.698*(12)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>betweenness</td>
<td>0.579</td>
<td>0.720*</td>
<td>0.558</td>
</tr>
<tr>
<td>Clerkenwell</td>
<td>63</td>
<td>closeness</td>
<td>0.095(93)</td>
<td>0.837*(90)</td>
<td>0.819(99)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>betweenness</td>
<td>0.585</td>
<td>0.773*</td>
<td>0.695</td>
</tr>
<tr>
<td>South</td>
<td>87</td>
<td>closeness</td>
<td>0.175(93)</td>
<td>0.688(24)</td>
<td>0.741*(27)</td>
</tr>
<tr>
<td>Kensington</td>
<td></td>
<td>betweenness</td>
<td>0.645</td>
<td>0.629</td>
<td>0.649*</td>
</tr>
<tr>
<td>Knightsbridge</td>
<td>90</td>
<td>closeness</td>
<td>0.084(81)</td>
<td>0.692*(33)</td>
<td>0.642(27)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>betweenness</td>
<td>0.475</td>
<td>0.651*</td>
<td>0.580</td>
</tr>
</tbody>
</table>

Table 27: Adjusted R-square values for correlations between pedestrian flows and least sum of length, least sum of angle change and least number of turns analysis applied to closeness and betweenness measures. Best correlations are marked *. Numbers in round brackets indicate best radius in segments for closeness measures.

<table>
<thead>
<tr>
<th>Area name</th>
<th>Gates</th>
<th>Measure</th>
<th>Least length</th>
<th>Least angle change</th>
<th>Fewest turns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barnsbury</td>
<td>117</td>
<td>closeness</td>
<td>0.119(57)</td>
<td>0.719(18)*</td>
<td>0.701(12)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>betweenness</td>
<td>0.578</td>
<td>0.705*</td>
<td>0.566</td>
</tr>
<tr>
<td>Clerkenwell</td>
<td>63</td>
<td>closeness</td>
<td>0.061(102)</td>
<td>0.637(39)</td>
<td>0.624*(36)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>betweenness</td>
<td>0.430</td>
<td>0.544*</td>
<td>0.353</td>
</tr>
<tr>
<td>South</td>
<td>87</td>
<td>closeness</td>
<td>0.152(87)</td>
<td>0.523*(21)</td>
<td>0.502(27)</td>
</tr>
<tr>
<td>Kensington</td>
<td></td>
<td>betweenness</td>
<td>0.314</td>
<td>0.457</td>
<td>0.526*</td>
</tr>
<tr>
<td>Knightsbridge</td>
<td>90</td>
<td>closeness</td>
<td>0.111(81)</td>
<td>0.623*(63)</td>
<td>0.578*(63)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>betweenness</td>
<td>0.455</td>
<td>0.513*</td>
<td>0.516</td>
</tr>
</tbody>
</table>

to streets with unrestricted two-way flows, but for pedestrian movement the whole urban network was included (though observations made inside a college and a housing estate were not). The pattern of adjusted R-square values for vehicular movement is shown in Table 26 and for pedestrian movement in Table 27 with a total of 48 correlations and therefore 16 possible best correlations. Correlations will be in general negative (but see below) since movement should increase with less metric, angular or directional change.

The results give a consistent picture. In 11 out of 16 cases, (5 vehicular and 6 pedestrian) least angle correlations are best. In the remaining five cases fewest turns is best, but in each case only marginally better than least angle. In no case is a metrically based measure best, and in no case is a least angle measure worst. On average correlations based on least length measures are markedly lower than the other two. The pattern of correlation for the metric, least angle and fewest turns interpretations of the closeness measure with increasing radius are shown from left to right in Figure 279, showing that the superiority of the least angle model is more marked than in the tabulated results in two of the four cases. The weakly negative correlations at low radii for metric (meaning that it
Table 28: Adjusted R-squares from axial rad-3 integration

<table>
<thead>
<tr>
<th>Location</th>
<th>Vehicular</th>
<th>Pedestrian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barnsbury</td>
<td>0.765**</td>
<td>0.706</td>
</tr>
<tr>
<td>Clerkenwell</td>
<td>0.627</td>
<td>0.570</td>
</tr>
<tr>
<td>South Kensington</td>
<td>0.819**</td>
<td>0.514</td>
</tr>
<tr>
<td>Knightbridge</td>
<td>0.560</td>
<td>0.467</td>
</tr>
</tbody>
</table>

is positively correlated with movement) are due to the fact that higher connectivity for
a segment will both tend to produce more movement and have a higher sum of lengths,
while higher sums of length at the larger scale will mean greater distance from the rest of
the system and so less movement.

6. Discussion

How then are these results to be interpreted? From a cognitive point, it is clear that, unlike
previous results from axial maps, the differences in movement correlations with the differ-
ent definitions of distance cannot be network effects, since in each case the representation
of the street network and its graph are identical, and all that differs is the mathematical
interpretation by varying the concept of distance. The differences in correlation can then
only be due to differences in the degree to which the each mathematical interpretation
coincides with the interpretations made by individuals moving in the system. It is then
an unavoidable inference then people are reading the urban network in geometrical and
topological rather then metric terms. Although it is perfectly plausible that people try to
minimise distance, their concept of distance is, it seems, shaped more by the geometric
and topological properties of the network more than by an ability to calculate metric
distances. In general we might say that the structure of the graph governs network effects
on movement and how distance is defined in the graph governs cognitive choices.

These results have three implications. First, they show that it is the geometrical and
topological architecture of the large scale urban grid that is, as space syntax has always
argued, the most powerful shaper of urban movement patterns. These factors are not
currently represented in most of the models currently in use to predict urban movement.
The effect is that the design of movement systems takes not account of the primary factors
which shape urban movement. Clearly, this situation cannot continue.

Second, the results show that axial graphs in their present form are in most circum-
cstances a perfectly good approximation of the impact of spatial configuration on move-
ment. In two of the eight cases reported, the correlation between movement and radius-3
integration is better than any of the segment analyses, and in general the spread of $r^2$
values for axial graph mirrors the pattern of correlation for the new, more disaggregated
segment based measures.

Third, these results are the strongest demonstration to date that the architecture of
the street network, in both geometrical and topological sense, can be expected, through its
effect on movement flows, to influence the evolution of land use patterns and consequently
the whole pattern of life in the city. This most powerful feature of the urban system can
surely not continue to be sidelined in urban modelling, and the architectural effects of
the large scale street network discounted. These issues will be discussed in a forthcoming
paper.(Hillier 2005)
Network effects and psychological effects: a theory of urban movement

Literature


