145

Introducing random walk measures to space syntax

Dror Fidler
The Bartlett School of Architecture, UCL
d.fidler@ucl.ac.uk

Sean Hanna
The Bartlett School of Architecture, UCL
s.hanna@ucl.ac.uk

Abstract

We introduce Random Walk Closeness (RWC) to the space syntax computational paradigm. Random walks are stochastic processes in which an unbiased walker traverses a network purely based on his current location. Random walks have been used by space syntax in an agent based scenario, where results are simulation based. Here, the results are mathematical based, i.e. RWC is derived from access times, which are the average number of steps it takes to walk between locations. Results suggest an improvement in correlating pedestrian movement over Integration, RA and network Closeness.

Keywords

Random walk, access times, integration.

1. Introduction

The notion of the random walk has always been implied by space syntax measures such as integration and choice, but not fully explored explicitly. Yet it is a notion that is central to related work in other fields of network analysis and mathematics. This paper investigates the properties of random walks in the context of an urban graph, with the intention that this may usefully bridge these fields and contribute to a better fundamental understanding of the space syntax measures.

The intuition of the random walk is similar to the graph structural properties measured in space syntax. In the case of integration, it is first a measure of paths. Integration is expressed formally as a property of the graph itself (i.e. depth) but its significance is typically described as a property of action within the graph, for example: as the "number of turns" required to move from node to node. Depth is used as an indication of accessibility. In the context of empirical observation, integration is usually presented against recorded data to demonstrate a correlation of higher integration values with higher density of pedestrians or vehicles observed in particular spaces. This again implies the association of the static graph with individual paths travelled.

Secondly, there is the assumption that these paths are random, capturing the notion of large group of varied people rather than a single individual. The key measure is mean depth, which is considered from the total of every other node and therefore assumes individual paths are not important in their details, but because these are independent and individual path goals randomly distributed across the population, the law of large numbers suggests that in the overall sum of paths the distribution of people will reliably converge. If space syntax is occasionally criticized as succumbing to spatial determinism, it is a misunderstanding of this difference between individual variation and collective reliability that is at fault.

A particular type of random walk is intentionally used in space syntax in visual agent simulations of space (Turner and Penn, 2002), in which each agent's movement is driven by a random selection of direction at each time step. It is clear that such agent paths are not individually representative of human movement, but that the aggregate distribution of many agents in space is. It is this aggregate behavior that is reliably influenced by the space represented by the graph that is considered via the space syntax measures of integration and choice to be a property of that graph.

Other, related fields, including mathematical graph theory and more applied network engineering, explicitly use the calculation of random walks. These also can be shown to converge on a steady state distribution over time; however this distribution can be determined analytically without the need for iterative simulation. The axial map and related representations have been used by mathematicians such as Blanchard and Volchenkov (2009), who refer specifically to space syntax and integration in their analysis of urban graphs, but even in these cases the fundamental method of analysis is not integration or choice, but the random walk.

These two measures are related to the centralities measured in pure graph and network theory: integration approximates closeness centrality, and choice approximates betweenness centrality. They are close analogues, but different from definition of those, particularly as they are refined for specifics of urban graphs dealt with in space syntax research. They are used in practice in part because of the wealth of empirical observations that they have been shown to explain, and for this reason, any alternative measure may usefully be compared to the same empirical observations. This paper reviews a calculation of centrality based on the random walk, Random Walk Closeness (RWC), and compares it with integration in two urban case studies well known to the community. Results from these suggest that RWC correlates better with observation than integration.

The following sections introduce the measure of Random Walk Closeness via a mathematical description of closeness and first passage time, before reviewing the results of the case studies.

2. Closeness in a graph

A graph consists of a node (or vertex) set often notated by V, and an edge (or arc) set notated by E, which is a binary relation on V. If E is a symmetric relation then G is called an undirected graph. A network centrality is a function defined on V which assigns importance to nodes according to certain criteria.

Closeness is one of the earliest network centralities, introduced by Bavelas (1950), and later refined by Beauchamp (1965). Given a connected network, represented as an undirected graph G = (V, E), with no self loops or multiple edges, the Closeness of a node V is defined as:

$$C_c(v) = \frac{1}{\sum_{u \in V} d(v, u)}$$

Where d(v, u) is the length of the shortest path from node v to node u.

Closeness, thus defined, is used in the calculation of Integration, and RWC also averages values over all possible steps it takes to reach a node. In integration this is treated as a direct measure of the graph; in RWC as a means to determine probabilities of random movement occurring at a node.

3. From graphs to Markov chains

The underlying random walk based on a given graph (such as the axial or segment graphs), is most easily represented using an object called a Markov chain. Its construction for a particular graph is quite simple, as illustrated in Figure 1. A star of 4 nodes is transformed into a Markov chain by duplicating each edge bidirectionally and assigning edge weights. The edge weights are assigned uniformly based on the degree of each node. The weights represent the probabilities of a walker moving to adjacent nodes from a given node, and are called transition probabilities. Clearly, in the case of a star the transition probability from each leaf (node with degree 1) to the center is 1. Therefore the average walking time (number of steps) from a leaf to the center is 1. But what of the opposite? What is the average time it would take to walk from the center, u, to a leaf v? We may represent this time in an equation. Denote the time H(u ,v), then in the case of the star in Figure 1:

$H(u,v)=1/3\times1+1/3\times(H(u,v)+2)+1/3\times(H(u,v)+2)=1/3+2/3\times(H(u,v)+2)$

The term $(H(u,v)+2)$ means that if we walk to another leaf other than v then we are sure to return to the center at the next step, meaning that 2 is added to the overall time. This reflects an important property of Markov chains, that they are memoryless, i.e. if the walk returns to a node after a while then the probabilities from that node are identical regardless of the time that has past.

Therefore from the equation above, $1/3\ H(u,v)=5/3$, giving $H(u,v)=5$.

So walking from a leaf to the center would always take 1 step, while walking from the center to any specific leaf would take an average of 5 steps.

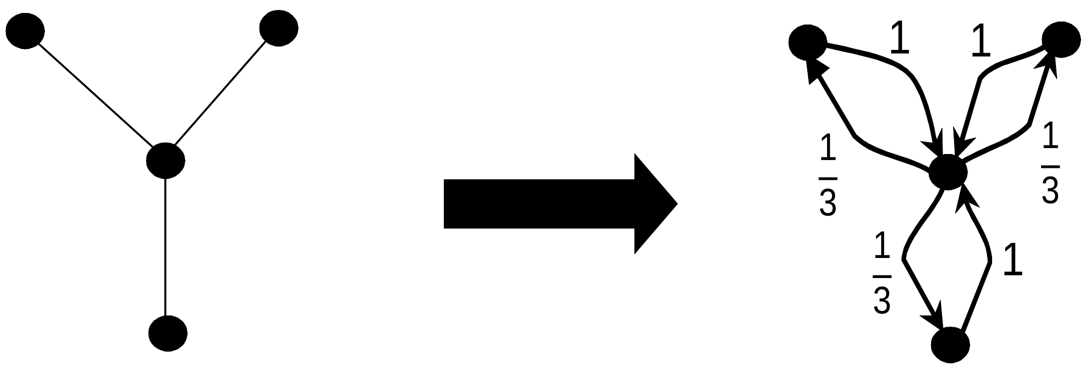


Figure 1: A star graph of 4 nodes transformed to a Markov chain.

4. First passage times by spectral decomposition

Real networks are far larger and complex than the star example given. However, there is a way to accurately calculate the first passage times, H(u, v), for any pair of nodes u, v. This is based on the spectral decomposition of a certain matrix called a Laplacian, which is based on the transition matrix of the Markov chain of a given network. In essence the calculation is similar to what was shown for the star, but it is easy to see that for a larger and more complex network the calculation would not be as simple, however There is a formula for H(u, v) that uses the eigenvectors and eigenvalues of the Laplacian matrix mentioned above.

5. First passage times and RWC

Formally speaking, denote by d(v) the node degree of v, and by π(v) the stationary distribution value for v, i.e. $\pi(v)=(d(v))/(2|E|)$. Let M be the Markov transition matrix on a graph G, and let D be a diagonal matrix such that

$D_{(i,i)}=1/(d(i))$. We define the normalized Laplacian T as :

$T=D^{-1/2}MD^{1/2}$

Where M is the standard random walk transition matrix on G.

Let $\{\Psi_1 \ldots \Psi_N\}$ be the set of real eigenvectors of T, and let $\mu_1 \ldots \mu_N$ be the real eigenvalues of T. The mean first passage time from node i to node j is the average number of steps needed to reach node j when starting from node i, and is given by:

$$H(i,j)=\sum_{s=2}^{N}\frac{1}{1-\mu_s}\left(\frac{\Psi_{s,j}^2}{\pi_j}-\frac{\Psi_{s,i}\Psi_{s,j}}{\sqrt{\pi_i\pi_j}}\right)$$

For a detailed overview and proof see (Lovász, 1993).

Random Walk Closeness (RWC) was introduced in (Noh JD, 2004). RWC uses the mean first passage time to a node to indicate its importance.

For a node v, the RWC value is defined as:

$$C_{RWC}(v)=\frac{N}{\sum_{u\in V} H(u,v)}$$

Note that the mean first passage time in this context is generally equivalent to the notion of mean depth in integration, but distances are not symmetric. Access times enable the analysis of the relation between two places, not based solely on network distance, but on the average time it takes to reach one place from the other. In mean depth, the shortest path is assumed, whereas in mean first passage time, it is a random path. This enables understanding the relation of one place to another asymmetrically while taking in to account the total network effects. The downside is that RWC takes longer to compute, but once the access time matrix is computed the data is available for all kinds of analysis.

6. Case study: Barnsbury and Kensington

The measure of RWC was compared to integration in two case studies, both to understand how it differs in general from integration and to determine its correlation to actual observation. Due to the longer computation time of RWC, the decision was taken in this first investigation to use axial graphs as opposed to segment graphs, and therefore to focus on integration as a comparison. The two London neighbourhoods of Barnsbury and Kensington were selected for ease of comparison with prior studies (Hillier B, 2005), from which sample data for pedestrian and vehicular movements associated with axial lines is available. This data consists of axial maps, and MPH (movements per hour) at certain sample locations (gates). The actual MPH values assigned to an axial line were the average MPH on the gates that were linked to each line in the original data.

We first compare the distribution of values of the RWC and Integration measures themselves, by plotting histograms of these for each of the two neighbourhoods (Figure 2). In this, we would expect the primary difference between the notion graph depth and that of a random walk first passage time to be evident. As noted for the star graph above, these times are asymmetric, resulting in a much longer distance from the centre to a perimeter node (5 steps) than from the perimeter to centre. Compared to integration, in which distances themselves are equal, we may expect to see a much lower range of values for the more peripheral nodes that make up much of the background of the city graph, as compared with the more central arteries. The histograms show exactly this, with a distribution of RWC skewed toward the less central, and a much thinner tail of more central nodes, whereas integration generally follows a more symmetric distribution.

Apart from this change of distribution, however, the axial maps themselves display a remarkable similarity between RWC and integration. An analysis of Barnsbury is plotted to include an ample surrounding boundary including Regent's Park to Kingsland Road, and Kensington including all the neigbouhoods surrounding Hyde Park, to south of the river Thames. In both cases, the rank of particular streets is similar under both measures, which is most clearly seen in the axial lines identified as highest integration or RWC—the same in both cases.

The maps look different, but primarily due to the overall distribution of values, rather than specific differences between streets. As noted above, the different distribution results in graphs of RWC with only a few lines in the high range indicated by red/orange in the plot, and most of the map in the much lower, blue values. The distinction between few axial lines in the highest range is more pronounced. While the more symmetrical distribution of integration is better suited to a regular, uniform, colour spectrum, plotting could be made clearer by changing the mapping of this colour range.

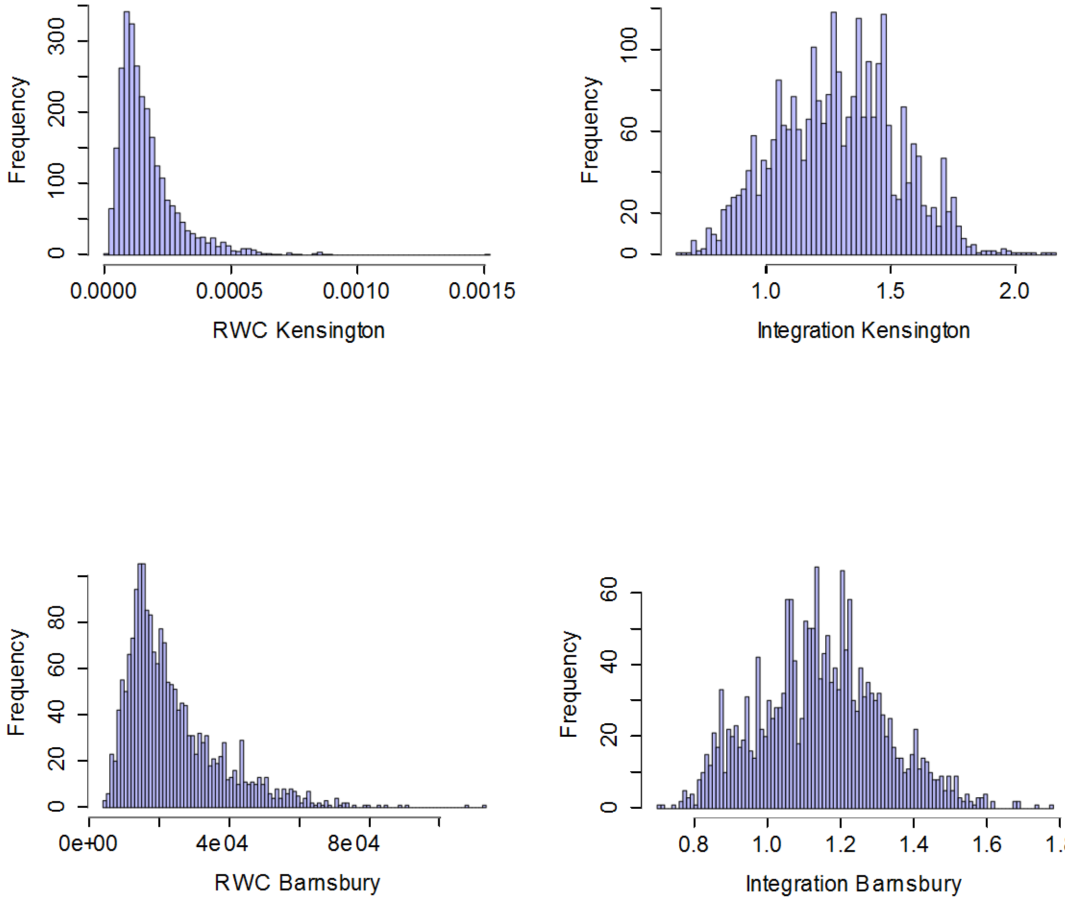


Figure 2: Histograms of the various measures.

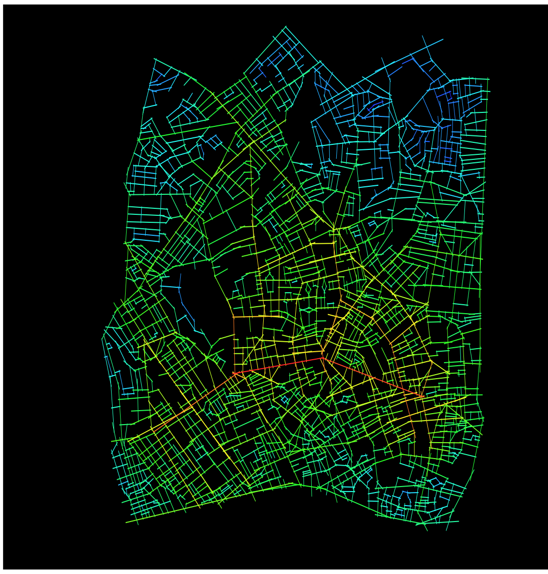
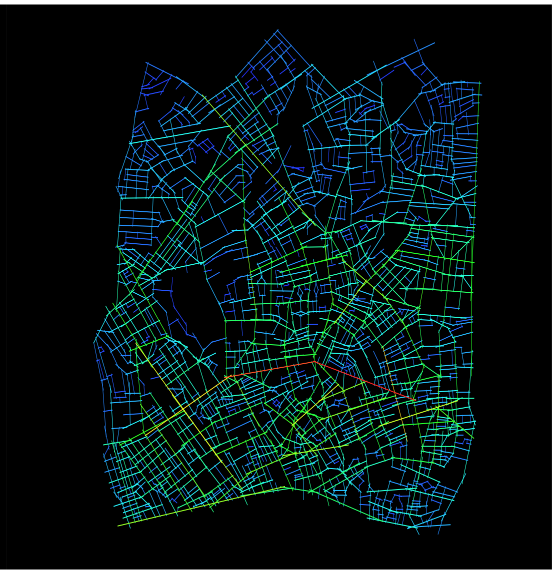



Figure 3: Integration in the axial map of Barnsbury. **Figure 4:** RWC in the axial map of Barnsbury

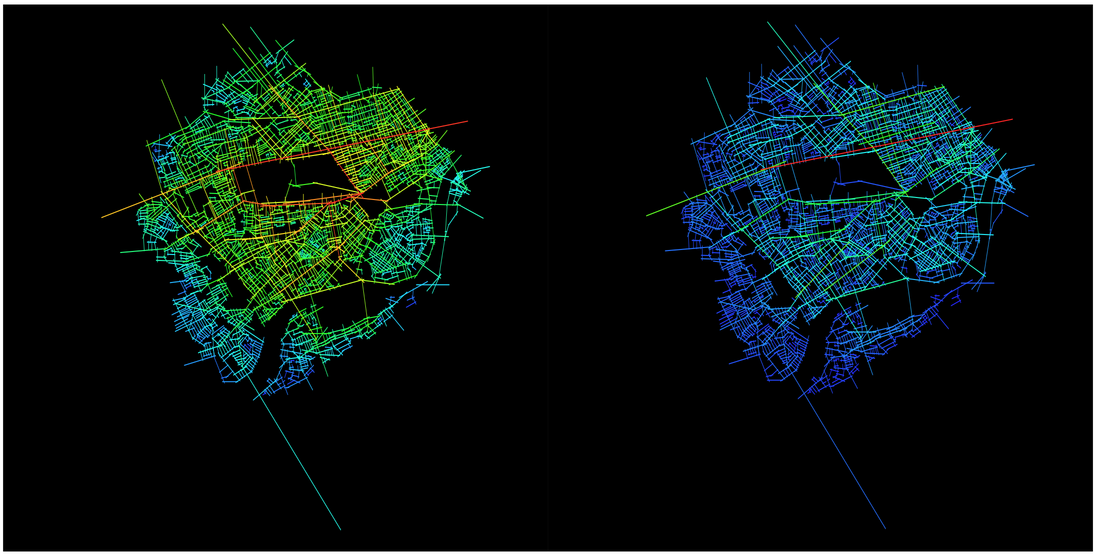
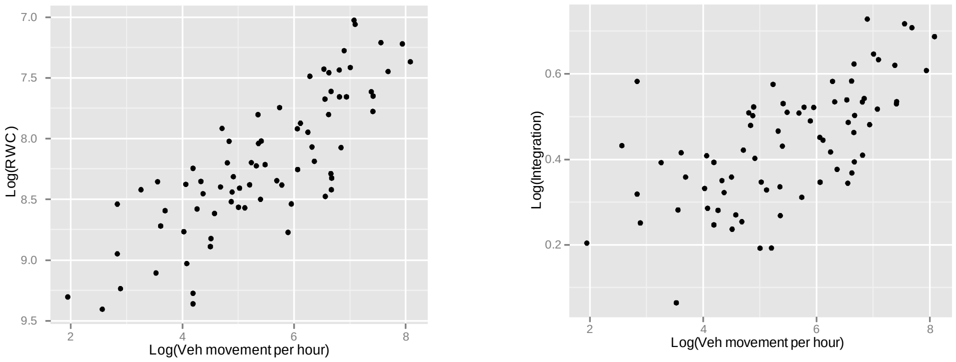


Figure 5: Integration in the axial map of Kensington. **Figure 6:** RWC in the axial map of Kensington

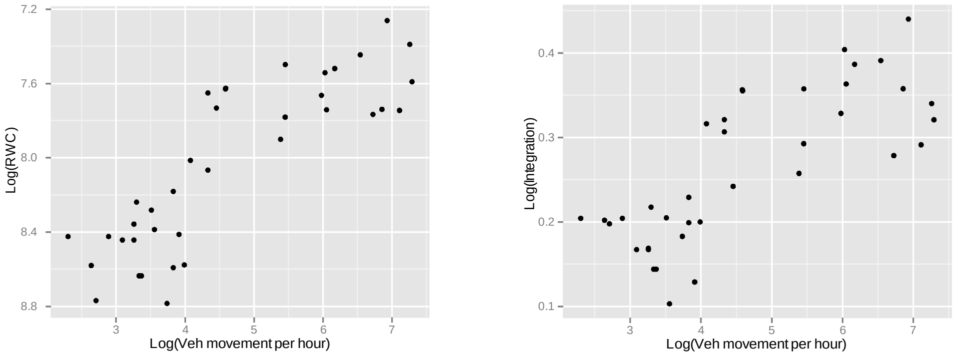
Of still greater importance than the relationship of RWC to integration is that to what integration explains. When the RWC measures are used as a prediction of actual movement, they appear to correlate better than integration to both pedestrian and vehicular data. Table 1. lists the Pearson correlation scores between both RWC and integration with observed pedestrian and vehicle movements in Barnsbury and Kensington. Both the raw correlations and those on a log/log scale are shown. In all cases there is significant improvement, with a mean of 20% better correlations (log/log) for RWC in both neighbourhoods (mean R=0.663 for integration, 0.797 for RWC). Scatter plots for each are shown in Figures 7 and 8.

Map	Type	RWC	Integration
Barnsbury	Ped log/log	0.842	0.669
Barnsbury	Ped raw	0.770	0.658
Kensington	Ped log/log	0.686	0.556
Kensington	Ped raw	0.654	0.542
Barnsbury	Veh log/log	0.845	0.775
Barnsbury	Veh raw	0.687	0.578
Kensington	Veh log/log	0.816	0.650
Kensington	Veh raw	0.747	0.664

Table 1: Pearson correlation scores for pedestrian and vehicle movements in Barnsbury and Kensington

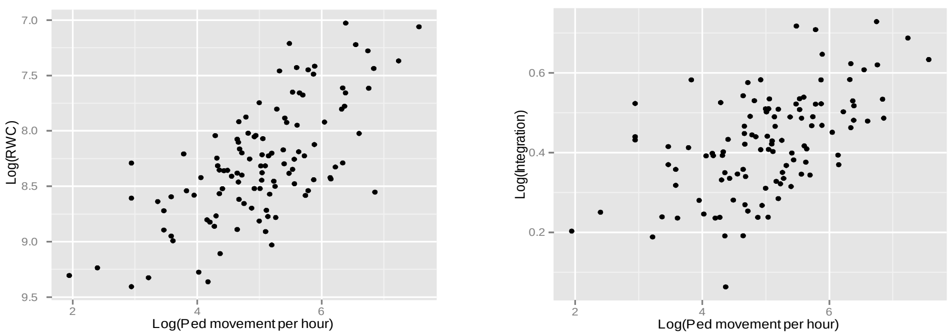


Kensington

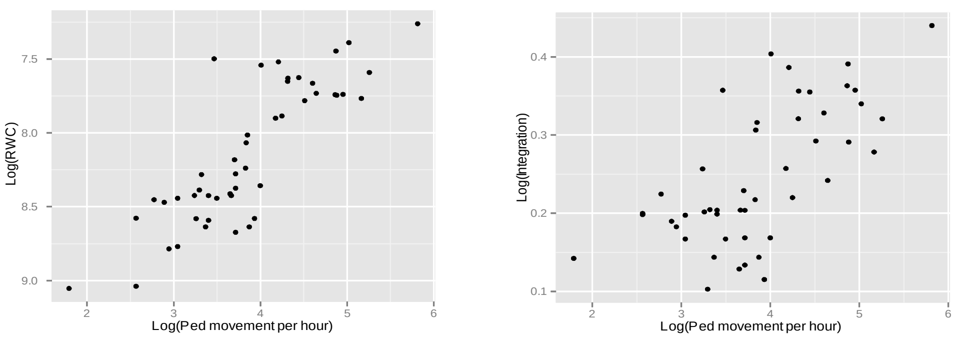


Barnsbury

Figure 7: Correlations of pedestrian movements to RWC and Integration



Kensington



Barnsbury

Figure 8: Correlations of vehicle movements to RWC and Integration

7. Discussion

There are several notable differences between the measures of integration and RWC, both with respect to what is measured and in the algorithm to do so. Because of the significantly greater probability in a random walk to frequent the better-connected nodes, and therefore that first passage time between adjacent nodes may differ more than mean depth, the results of RWC should be expected to be less continuous from one node to the next. When a travelling agent runs into a high degree node, this will impact the subsequent walk.

This is related to the asymmetry mentioned above. Where depth between two nodes in a graph is the same from x to y and from y to x, the random walk means a more central node appears 'closer' from a peripheral node than that same peripheral node appears from the central. A particular place might look more isolated from x than from y. The details of this can be easily explored in the hitting time matrix, which may reveal big discontinuities for particular places in the graph. Making this matrix available might yield a very useful additional tool for analysis.

The computational complexity of the algorithms to calculate the measures differs, potentially making RWC less practical for calculation than integration in come conditions. Let $|V|=N$. Integration requires the distances between nodes. The distances to a specific node can be computed in $O(|E|\log(N))$, and since the graphs we are dealing with are sparse, that is $O(N\log(N))$. For the entire graph that entails a theoretical computational complexity of $O(N^2 \log(N))$, although in practice integration is calculated by a flow algorithm that is faster than this. RWC, on the other hand, requires the spectral decomposition of the laplacian matrix which is a computationally heavy task. This means we have to find the N eigenvectors and eigenvalues which has a time complexity of $O(N^3)$.

This paper has focused on the comparison of integration and RWC in axial maps, but demonstrates several other immediate possibilities. First, RWC is not the only existing random walk measure; a random walk version of betweeness centrality exists which may show analogous improvements to choice. Second, for comparison with the existing case study for which sampled data corresponded to axial lines, RWC was applied to an axial, not a segment map. The computation can equally be applied in its existing form to a segment map or any other graph representation, with the obvious issue of increased computation time of segment over axial representations due to the higher number of nodes. Furthermore, it may be possible to compute fewer eigenvalues and eigenvectors while closely approximating RWC, thus reducing time complexity. This would require both a mathematical proof of the accuracy of such approximations and empirical evidence of their performance.

8. Conclusion

Random walk measures are related in general principle to the phenomena explained by existing space syntax measures, and are supported by significant research both theoretical (mathematics) and applied (engineering), but less so in the context of urban and other spatial networks. The wealth of observed data in these domains has to date primarily been described by the space syntax measures. This study has aimed to bridge the two. Comparing integration with RWC indicates strong similarities between the two measures, but the analysis of actual data suggests that RWC is a better approximation to observed pedestrian and vehicular movement.

For the overall picture of the city given by the two measures, there is a rough equivalence in results: there is a similar ranking of more and less central street lines, as seen in the plots of the axial maps by integration and RWC values, albeit with different overall distribution. Many of the conclusions that would be drawn, such as the location of the city core, the deformed wheel pattern, etc. would be identical, and this equivalence could potentially put a great deal of previous spatial research on a firmer footing with respect to mathematics and related theory.

The use of RWC, as proposed here, does appear to result in improved correlation to the empirical data, so it is potentially a better way of explaining such observations in future research. If this continues to hold, the notion of the random walk itself, as a conceptual can descriptive tool, may help to provide a better means to understand and explain what occurs in the natural movement through space.

Acknowledgements

This paper was produced from research funded from the EPSRC Platform grant awarded to the Space Group at the Bartlett Faculty of the Built Environment, University College London (grant reference EP/G02619X/1).

References

Bavelas, A. (1950), 'Communication patterns in task-oriented groups'. In J. Acoust. Soc. Am, 22(6), pp. 725–73.

Beauchamp, Murray A. (1965), 'An improved index of centrality'. In Behavioural Science, 10(2), pp. 1099-1743.

Blanchard, P and Volchenkov, D. (2009), Mathematical Analysis of Urban Spatial Networks, Springer-Verlag, Berlin.

Hillier, B. (2005), 'Network and psychological effects in urban movement'. In Proceedings of Spatial Information Theory, Ellicottsville, N.Y.: Springer-Verlag: Berlin, Germany, pp. 475-490.

Lovász, L. (1993), 'Random walks on graphs: a survey'. In Combinatorics, Paul Erdös is Eighty (Volume 2), pp. 1-46.

Noh, J.D, R. H. (2004), 'Random walks on complex networks'. In Phys Rev Lett.

Turner, A and Penn, A. (2002), 'Encoding natural movement as an agent-based system: an investigation into human pedestrian behaviour in the built environment'. In Environment and Planning B: Planning and Design, vol. 29 no. 4, pp. 473-490.