431-2

RANDOM WALKS ON WEIGHTED GRAPHS

CENTRALITY MEASURES OR SIMPLE AGENTS

SEAN HANNA

ABSTRACT

Two new methods of graph segment analysis are introduced and tested, which use a random walk as a theoretical and practical basis. The first, random walk betweenness (RWB), is a measure of betweenness centrality, similar to choice, and the second is a simple calculation of a random walk on a graph with similarities to visual agent simulation. To implement these, two methodological innovations are made: angular weighting for probability of turns and a dual node representation of segments. The two methods are tested for their correlation with both pedestrian and vehicular movement and compared with segment angular choice. RWB is seen not to be as effective a prediction of human movement, but simple random walks appear approximately as effective as choice, and faster to compute. Practically, this yields a substantially faster algorithm for movement prediction than previously exists. Theoretically, it suggests a much simpler model of agency: that navigation may be more opportunistic than optimised. To the extent the latter is valid it may prove valuable to a more complete theory of movement and a bridge between Space Syntax and related fields of mathematical graph theory.

KEYWORDS

Random walk, choice, agents

1. INTRODUCTION

The notion of random movement in graphs underpins Space Syntax accounts of behaviour in space. Two separate domains of analysis have developed, differing in whether this random movement is merely implicit, or used explicitly in the model. Graph representations of space, including axial (Turner et al. 2005), segment (Hillier and Iida, 2005) or visibility (Turner et al., 2001) are most typically assessed directly by measures of graph centrality (choice, integration), without explicit acknowledgement of the movement within them, but by describing aggregate movement without any specifics of individual difference or agency, the implication is that these are random. In visual agent simulation (Turner and Penn, 2002) random movement is modelled explicitly, by agents repeatedly choosing their direction randomly from a weighted probability distribution based on their isovist.

Work in related fields such as mathematical graph theory and network engineering (Blanchard and Volchenkov, 2009) use the random walk as a theoretical basis because its statistical properties are known and well researched. Previous work (Fidler and Hanna, 2015) has attempted to bridge the fields of mathematical graph analysis and Space Syntax by comparing a measure of centrality based on the random walk, random walk centrality (RWC), with integration on axial maps, and have found RWC to correlate better with observed movement. To date, none of these methods have applied to graph segment representations, although segment representations have several advantages over axial lines. Notably, the analysis of segments includes an angular weighting between each segment, a property which has been shown to be fundamental to human movement, when compared to metric distance, for example (Turner, 2009).

This property of angular weighting provides a potential link between the two Syntax analysis domains in that it suggests a potential probability distribution in which random movement can be interpreted on such graphs. Although angular weighting is interpreted as a continuous cost function that applies to

routes through a graph, the weights themselves exist between individual nodes. Two perfectly straight segments (0 degrees) have a cost of zero, and turns of increasing angle have greater cost: 90 degrees a cost of 1.0, up to a maximum (180 degree) of 2.0. But cost is notionally the inverse of the probability of choosing a destination. If the angle between segments correlates more directly with the probability of it being selected by a moving pedestrian or driver, it is not immediately clear what this distribution is. Recent work (Frith 2017) suggest that even cost might not be a simple linear function. This can be addressed as an empirical question, and is so done here by testing different functions of angular weighting.

This paper introduces and tests two different approaches to random walk analysis of the angular weighted segment map. The first is a calculation of random walk betweenness (RWB) centrality, as developed by Newman (2005), a measure most closely analogous to the Space Syntax measure of choice, but which differs in considering a probabilistic distribution of possible (angular) shortest paths between source and destination, rather than a singular shortest path. The second is a simple calculation of the effect of random movement over the graph over time, with the probability of choosing adjacent segments given by a function of their angle. Both differ from choice in that angles contribute directly to the probability of movement instead of via a cumulative cost function, and various issues of implementing the calculation that arise from this are presented. In addressing these, the paper introduces two methodological innovations: an angular weighting function indicating likelihood as opposed to cost, and a dual node representation of segments.

Random walk betweenness does not predict movement as well as certain radii of angular choice, but the far simpler calculation of random movement, weighted by angular change, is comparable or superior to the best existing measures at predicting both pedestrians and vehicles. This appears to be a significant methodological contribution as the calculation is also far more efficient, and therefore quicker, than either random walk betweenness or shortest path betweenness (choice). Furthermore, it introduces a new connection between the related fields of Space Syntax and mathematical graph theory.

2. METHOD

The two methods are tested in case studies of London neighbourhoods: Barnsbury, Clerkenwell and Kensington, where they are compared with various existing measures of angular choice and with observed pedestrian and vehicular movement. For continuity with previous research, these are the same cases used in previous tests indicating the validity of random walk based centrality measures for axial maps (Fidler and Hanna, 2015), initial Syntax proposals of betweenness centrality or choice (Hillier and Iida, 2005) and previous studies of urban movement (Penn et al., 1998).

Angular weighting

Angular weighting between nodes (segments) in the graph captures the known human preference for straighter paths over angled. The cost function used in the angular weighting of paths is a linear mapping of the angle between segments (0 to 180 degrees) to a range 0.0 to 2.0, such that the greatest changes of turn have the greatest value. A probability distribution is the reverse: we require the relative likelihood of choosing a given turn at each location, so the least angled changes of segment will have the greatest value. For a range of [0,1] it is reasonable to identify the maximum value (1.0) with straight ahead (0 degrees, or 0 angular weighting cost function) and values approaching 0 probability with greatest turns (180 degrees, or angular weight of 2.0). What is not clear is how intermediate values should be distributed within that range, as cost and relative likelihood are not necessarily linearly related (Frith, 2017).

A reversal of the simple linear weighting of angles (180 to 0) to the range [0,1] yields a 0.5 value for right angled, 90 degree turns. But is a 90 degree turn 1/2 as likely as continuing straight as this suggests, or 1/20 as likely? Raising this range to various exponents yields a variety of different non-linear functions of angles between the minimum and maximum. Raised to the power of 2, a right angled turn is $25\% (1/2^2)$ as likely to be taken as a straight path, to the power of 5, it is only 3% as likely $(1/2^5)$. This is tested empirically by raising the values of this range to successive exponents, to determine whether any particular distribution correlates best with actual movement.

Random walk betweenness

Newman (2003) proposed a calculation of random walk betweenness (RWB) as an alternative to betweenness measures based on shortest paths. These previous measures of betweenness centrality, like *Choice* in Space Syntax, calculate betweenness of node *n* as a proportion of all shortest paths between other nodes that pass through node n. Where more paths pass through n, betweenness centrality is higher. In the case of angular choice, a shortest path is defined as the least angular distance (turns) taken along that whole path. The assumption implicit in these definitions is that paths taken, e.g. by pedestrians in a city in *Choice* analysis, consist only of these optimal short paths. RWB "relaxes this assumption": while still preferentially weighting shorter paths, it allows for alternative routes through the graph to be considered. Rather than considering only one potential path from nodes a to b, RWB calculates a probability distribution for all possible routes, and uses this to calculate the betweenness of any given node n. Although some nonsensical routes could be imagined in which a node is passed multiple times by doubling back on the path, these implausible routes are eliminated from the count by considering only the net traffic through a node. Compared with a count only of optimal paths this account of multiple paths would appear more a more realistic description of pedestrians who do not have perfect knowledge of a city. Newman's (2005) method is followed in the calculation of RWB, with a weighting between graph nodes provided by a function of the angular change as described above. Calculation is made using any nodes a and b across the whole graph, without restriction of their distance, for comparison with choice radius n. Several further considerations are given to the handling of the graph representation, as follows.

Like *choice*, RWB is computationally expensive, scaling with the number of nodes in the graph (n) on the order of $O(n^4)$. In practice, the estimation of choice can be done by calculating only a random sample of nodes between which to calculate shortest paths, and the same procedure is applied here. For a graph of n nodes, a smaller subset (m) is selected from which to calculate all path origins and targets, while RWB is calculated for all nodes. This yields a run time of $O(m^2n^2)$, which is identical for m=n, but much more efficient for smaller m. In the case of the graphs used here, where maps of approximately 1000 nodes are used (after pruning as below), a subset m of 80 to 100 was found to yield a good approximation of RWB (variances comfortably <5%), so this was used for faster calculation in the experiments presented here.

The maps used in the three case studies are also cropped for efficiency of calculation. Each original map covers an area of approximately 5km across, however, traffic data is given only for a smaller region of approximately 100 nodes in the centre, and it is this central area only that is used to assess correlation with movement in this paper. For RWB calculations, the three maps were cropped to a circular radius of 2 km diameter, each of which fully contains the region for which traffic data are used. This potentially introduces an edge effect whereby the removal of peripheral nodes changes the results of RWB values for nodes in this region, particularly as the analysis is not limited to small radii. In this case, the difference in correlation values for the un-cropped map between choice of higher radii (>1 km) and radius n had previously been seen to be low, so the risk of the edge effect was deemed acceptable, but may be worth further investigation if RWB warrants future research in this domain.

The variation in street segment length was also considered in the graph representation for RWB. The angular cost function used in the calculation of choice is considered only with a total sum of angular change in a route, regardless of the number of steps (street segments) required to make the journey. This does not translate directly to a scenario in which random walks are considered, because there is a distance inherent in each step of a random walk. In the case of a route along a straight street consisting of a number of segments (joined at 0 degrees), for example, a measure of angular distance as used in *Choice* would yield the same result, zero, regardless of whether the route covered three segments or thirteen. In the case of a random walk, even one in which the probability is maximised (1.0) to step to each adjoining segment along the straight street, there is a distinct difference between a walk of three steps compared to one of thirteen. In calculations of RWB, there would therefore be a bias toward streets with longer segments, as the same distance would inherently appear shorter. This was mediated in the graphs by splitting longer segments into two or more nodes, with a maximum connection weighting of 1.0, equivalent to 0 degrees angle.

Random walks

The direct simulation of a random walk is far simpler than RWB. The graph is defined by a square $n \times n$ matrix A, in which A(i,j) is the value of the angular weight (0.0 to 1.0, as described above)

between any connected nodes i and j, or A(i,j)=0 for all unconnected nodes. A distribution of walkers (agents, pedestrians) on that graph is an n×1 column vector S. A transition matrix T is created by normalising all columns in A such that the sum of all values in that column equals 1.0, indicating the total probability of a walker moving to any of the possible adjacent nodes and conserving the number of walkers on the graph at any time. The new distribution S' of agents on the graph is calculated simply by multiplying the previous distribution by the transition matrix: S' = S × T.

In the analyses presented here, the initial distribution is a column vector S of ones, indicating a single agent on each node of the graph. Subsequent steps were calculated until the distribution was seen to converge on a steady state: 100 steps for all examples in this paper.

A crucial innovation was made in the basic graph representation, which uses two nodes to represent each street segment, rather than one, with each node representing a different direction of travel. This was necessary to reflect the nature and cost of 180 degree turns. In the case of a graph in which a single node is used for a segment, the angular weighting to connecting segments at each end of that segment would be identical. In the case of a random walk, a walker is therefore just as likely to double back by making a 180 degree turn onto a segment from which they have just arrived as they are to continue straight ahead, or just as likely to make a 135 degree turn as a 45 degree turn. In calculations of angular *choice* such paths are not considered because they are be definition not the shortest, and in RWB the calculation of net passage rates ignores cases of doubling back on a node. For a straight random walk, however, the representation of each potential direction as a separate node in the graph eliminates this possibility.

The dual node representation is relevant to space syntax visual agents and visibility graph analysis (VGA). For grid based representations of space used in such analyses, where a node represents a grid square, graph connections exist to all other nodes within that isovist. Agents make random walks within such graphs, as they do in this paper, but apart from experiments with "through vision" (Turner 2007), the agent's possible directions are limited to a 170 degree field of view, matching human vision. The direction of this vision is determined by the direction from which the agent has entered the spatial position therefore has a different set of next step probabilities for each possible orientation of the agent, and is correctly represented by a different graph node for each possible orientation. The same is true of the random walk analysis presented here, only the graph is far simpler as only two directions are possible for each node.

3. RESULTS

The results of both RWB and random walks were checked for correlation with actual pedestrian and vehicle movement data taken from previous studies published in (Hillier 2005, Fidler and Hanna 2015) so as to be comparable with existing work. The degree of correlation was compared to that of radius n angular choice of the same neighbourhood, which is taken as a baseline, or current state of the art for analysis of such urban street segment maps. While an initial review of existing measures across all three case studies revealed that radii other than radius n were occasionally better predictors of movement in certain cases, a posteriori, all the highest correlations tended to be with larger radii, and radius n was considered to be a priori the most reliable predictor of movement in general. Figure 1 shows the radius n choice values for all segments in the Bloomsbury map, and the actual pedestrian counts of the smaller sample of segments from the centre of this map. A broad picture of traffic and its prediction by choice is evident visually, with the longer, straighter routes of both maps showing high values of choice and high pedestrian counts respectively. There is a good correlation between choice and observed traffic, with a Pearson coefficient of 0.88 (log/log 0.82) for pedestrians, and 0.64 (log/log 0.85) for vehicles.

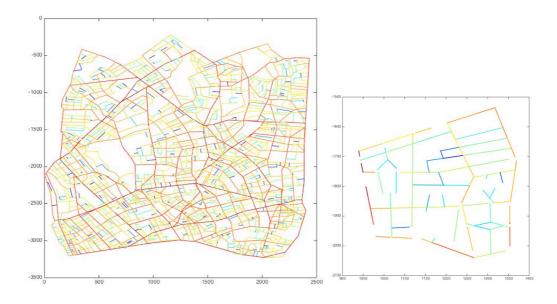


Fig 1: Left: A baseline: choice radius n for Barnsbury and surroundings, with colour values scaled logarithmically from dark blue (lowest) to red (higest). Right: actual pedestrian movement in the sample area (centre of left map), with colours scaled linearly.

RWB

The results of RWB for the same map of Bloomsbury are shown in figure 2, using two different functions of angular weighting between segments: a linear function of angle, in which, for example, a 90 degree turn is half as likely as an option to continue straight, and a function of angle raised to the power of 4, in which the 90 degree turn is only 1/16 as likely. Visually it is evident that the second, exponential function more closely resembles the maps of choice and pedestrian movement in fig 1, with higher values of RWB occurring along the longer straighter routes in the map.

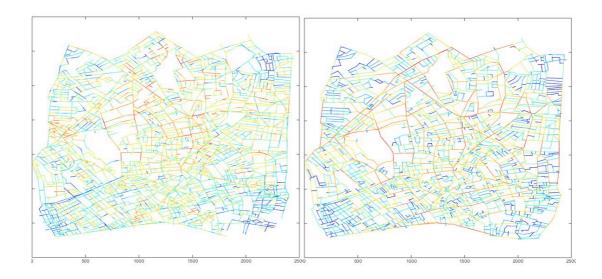


Fig 2: RWB for Barnsbury and surroundings. The left map uses a transition probability in proportion to the angular weight, the right uses angular weight ^4. Colour values for RWB are logarithmically scaled.

The question of which function of angle allows RWB to most closely approximate real human movement is addressed by examining the correlations with traffic using a range of exponents. Figure 3 plots these for both the Barnsbury and Kensington maps. In all cases, the basic linear function of angle is not a good predictor of movement, but as the exponent rises correlation increases to a peak, and then decreases. The overall trend is the same across all observations, but this peak differs

markedly depending on whether the correlation is taken of the actual values of RWB and traffic, or of the logarithm of these values. In the case of the vehicular correlations in Barnsbury (left) the log correlations peak at angle³ or angle⁴, whereas the correlations of actual values increase approximately linearly with the angle exponent up until a peak at angle¹⁰ or angle¹¹. These are the most obvious version of a trend across all data sets that indicates the functions of angle correlating best with movement are have exponents in the range of 11 to 16 for actual values and of 2 to 8 for log values.

The overall range of Pearson coefficients are generally lower than those of choice, except in the peak (angle¹¹) case of the vehicular correlation in Barnsbury (r=0.87). Comparison of the effectiveness of each method in predicting traffic will follow after a review of random walks.

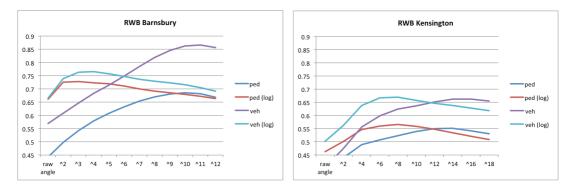


Figure 3: Correlation of RWB with pedestrian and vehicular movement using various exponential functions of angle. Correlations of actual values, and log/log correlations are shown.

Random walks

The direct simulation of random walks necessarily goes through a number of iterations, as described in the method section above. The successive iterations are a completely deterministic multiplication of probabilities, so always provide the same output, which would be equivalent to the mean of all possible random walks for that number of steps. The first state consists of an uninformative, uniform distribution (notionally of "agents" or traffic) equally over all segments, but should converge on a steady distribution over time, and so the intent as an analytical tool is that it is the steady state distribution that is of interest. Figure 4 shows a plot over 100 iterations of the Pearson correlation of the random walk with pedestrian movement in Barnsbury. As with tests of RWB, a series of successive exponents of the segment angle (from 1 to 5) are used as the relative transition probabilities between nodes. In all cases, there is indeed negligible change in any of the correlations after about 50 iterations, so the end state of the run of 100 iterations is taken as the final, steady state.

Two observations are relevant. The first is that, as with RWB, there is a steady increase in the degree of correlation as the exponent of angle increases, until a peak at angle⁴, after which it begins to decline. This is unsurprising given the same observation in RWB. The second is the correlation values as iterations of the random walk progress. These begin low (with the uniform distribution of traffic) but rather than rising gradually to the maximum correlation at the steady state, peak or near peak correlations are actually found quite rapidly, from between 12 and 18 iterations. These then either level out, or decrease slightly for the remainder of the walk. For higher exponents of angle, 4 and particularly 5, the peak is followed immediately by a drop before the values converge. The reason for this early peak suggests further research, but it seems plausible that it may be related to the finite nature of journeys, as real people do not continue walking or driving through the map indefinitely.

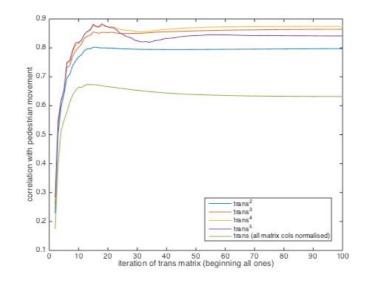


Figure 4: Correlations with pedestrian movement in successive iterations of the random walk.

To determine the most appropriate function of angle for use in random walk simulation, as with RWB, the correlation between both pedestrian and vehicle movement and the random walk distribution was examined for successive exponents. Figure 5 shows these for both Barnsbury and Kensington. The same plots of pedestrian and vehicle, actual and log values, are shown as in figure 3 for RWB, only because of the iterative nature of the walks, the correlations of both the final state (100 iterations) and the peak values (generally between 9 and 21 iterations) are shown. Although these also cover a range of values, the exponents of angle that best approximate observed movement are almost always in the range of angle³ to angle⁵ (Barnsbury pedestrians, with angle⁶ or angle⁷, being the outlier).

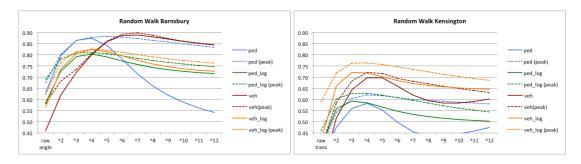


Figure 5: Random walks using different exponents of the angular weight. Peak value iterations are shown in dashed lines.

The transition function of angle⁴ produces the highest overall correlation of random walks with movement. This same function was similarly effective for many of the RWB analyses, particularly those using log/log correlations. The fact that there is some discrepancy in the most highly correlated exponents of angle in RWB suggests that it would be premature to conclude angle⁴ actually represents human movement as a function of angle, but it is a reasonable to use this as a practical default probability for transition in both the RWB and random walk methods. The effect of this is that the probability of selecting a 90 degree turn is only 1/16 as likely as selecting a route straight ahead. Routes with minimal changes of angle will therefore be quite highly biased over those with greater changes, although not excessively so.

Figure 6 shows the values of the distribution after a converged random walk (100 iterations, approximating a steady state) on the Bloomsbury map. Compared to figure 1 this shows a pattern very much like that of choice, and correlates similarly well with actual movement.

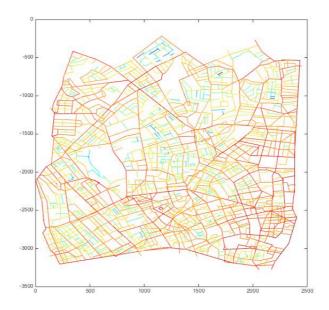


Fig 6: Converged random walk (100 steps) for Barnsbury and surroundings, using angular weight^4. Colour values for RWB are logarithmically scaled.

| Мар | Choice rad n (<i>log/log</i>) | RWB (log/log) | Walk 100 (log/log) | Walk peak (log/log) |
|------------------|------------------------------------|---------------|-----------------------|---------------------|
| Barnsbury (ped) | .88 (.82) | .58 (.61) | .87 (.80) | .88 (.81) |
| Barnsbury (veh) | .64 (.85) | .68 (.77) | .80 (.82) | .80 (.83) |
| | | | | |
| Chelsea (ped) | .78 (.71) | .59 (.35) | .37 (.37) | .60 (.51) |
| Chelsea (veh) | .79 (.83) | .50 (.47) | .41 (.41) | .50 (.60) |
| | | | | |
| Kensington (ped) | .51 (.66) | .49 (.55) | .58 (.59) | .62 (.63) |
| Kensington (veh) | .68 (.77) | .56 (.64) | .70 (.72) | .72 (.76) |
| | | | | |

Table 1: correlation between various analyses and observed movement for all three cases. Log/log correlations are shown in parentheses.

Comparison of Choice, RWB and Random Walks

Degrees of correlation of all three graph analyses, choice, RWB and random walks with observed movement are listed in Table 1, in which both new measures can be compared against the baseline of radius n choice (in bold) in terms of their predictive value of pedestrian movement. For all analyses with the two new methods, RWB and random walks, a transition probability function of angle⁴ is used, as described above.

Despite being related to choice in terms of measuring betweenness centrality RWB does not appear to be as effective a predictor of observed movement. Correlation values are almost all significantly lower.

Values for the much simpler random walk algorithm, however, are comparable to choice in most cases. Values for Barnsbury and Kensington are approximately equal or higher than those of choice, yielding good predictions of both pedestrian and vehicular movement in both the early iteration peak and in the final steady state. Values for Chelsea are lower, both in the steady state and peak, as they were in the RWB analysis. Whether there is a crucial difference between the Chelsea map and the others that would explain this remains a question that must be left for future research.

4. CONCLUSION

Two new methods of analysis were introduced and tested in this paper. The two methods of analysis are alike in that they both explicitly use the random walk as a theoretical and practical basis for measurement, but they yield different results and correlations with observed movement. RWB, a more complex and computationally costly algorithm similar in structure to that of choice, is not a particularly effective predictor of human movement. The more direct random walk, is a good predictor of human movement, to choice.

From a practical perspective, this effectiveness of the simple random walk simulation is significant in that it is very efficient computationally, far faster than both choice and RWB. Where the latter has a run time on the order of $O(n^4)$, or $O(m^2n^2)$ as described above, the random walk is only $O(n^2)$. Even with approximations applied to RWB to improve computation speed, this means a difference for the maps in this paper of seconds (for random walks) as opposed to tens of minutes (for RWB). This difference would be far greater for larger graphs, as n increases. If further research confirms the utility of the random walk measure, it is likely to be a very useful alternative for existing measures like choice.

In both methods, the angular weighting between nodes is crucial, but its effect is the reverse of that normally considered in Space Syntax. Where measures such as angular choice are interested in a cost function between nodes, where larger angles are more costly, the random walk is interested in a likelihood, where smaller angles are more likely. The precise function to map angle to likelihood was tested empirically, with the result that angle⁴ yielded the best overall approximation of actual human movement. This has the effect of a turn of around 28.5 degrees being half as likely to be chosen as a straight turn, and a right turn of 90 degrees being 6.25% as likely to be chosen.

Theoretically, the extent to which random walk methods are successful at predicting movement would imply the basic act of urban navigation may be much more locally opportunistic and undirected rather than goal directed or optimised. But of the two methods tested here, one is successful, and the other less so. Superficially, the result that basic random walks do correlate well with observed movement suggests a much simpler explanation for aggregate urban movement: that the destinations involved in, and route optimization implied by, measurement of choice are unnecessary. The random walk is a far simpler model of agency. However this appears in conflict with the RWB result, which suggests that to the extent that people do have a destination in mind, they are very good at selecting the optimal least angle path—choice consistently outperforms RWB. The most plausible explanation would seem to be that actual human movement, in which case it would appear that the random walk is not the whole, but must be an essential part of a complete theory of movement.

REFERENCES

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

Fidler, D., & Hanna, S. (2015). Introducing Random walk measures to Space Syntax. *Proceedings of the 10th Space Syntax Symposium*, 145:1-145:9. London: Space Syntax Laboratory, The Bartlett School of Architecture, University College London.

Frith, MJ. (2017). Using Qualitative Distance Metrics in Space Syntax and Configurational Analyses. Proceedings of the 11th Annual Space Syntax Symposium, pp. 161.01-161.15.

Hillier, B., and Iida S. (2005). Network and psychological effects in urban movement. In *Proceedings of Spatial Information Theory*, ed. A. G. Cohn and D. M. Mark, Ellicottsville, N.Y.: Springer-Verlag: Berlin, Germany, pp. 475-490.

Newman, MEJ. (2005) A measure of betweenness centrality based on random walks, *Social Networks* 27 (1), pp. 39-54.

Penn, A., Hillier, B., Banister, D., Xu, J. (1998) Configurational modelling of urban movement networks. *Environment and Planning B: Planning and Design*, 25 (1998) pp. 59–84.

Turner, A. (2007) To move through space: lines of vision and movement. In: Kubat, AS and Ertekin, O and Guney, YI and Eyuboglu, E, (eds.) *Proceedings of the 6th Annual Space Syntax Symposium*, pp. 037.001-037.012. Istanbul Technical University: Istanbul, Turkey.

Turner, A. (2009), The role of angularity in route choice: an analysis of motorcycle courier GPS traces, *COSIT'09 Proceedings of the 9th international conference on Spatial information theory*, pp. 489-504.

Turner, A. Doxa, M., O'Sullivan, D. and Penn, A. (2001). From isovists to visibility graphs: a methodology for the analysis of architectural space. *Environment and Planning B*. 28 (1), pp. 103–121.

Turner, A and Penn, A. (2002), Encoding natural movement as an agent-based system: an investigation into human pedestrian behaviour in the built environment. *Environment and Planning B: Planning and Design*, 29 (4), pp. 473-490.

Turner, A., Penn, A. and Hillier, B. (2005). An Algorithmic Definition of the Axial Map, *Environment and Planning B Planning and Design*, 32(3), pp. 425-444.