ON THE METRIC, SYNTACTIC AND FUNCTIONAL STRUCTURE OF AXIAL MAPS

008

Roy Wagner
Academic College of Tel-Aviv Yaffo

Keywords:

Urban modelling
Urban structure
Scaling
Random walk

Roy Wagner
Academic College of Tel-Aviv Yaffo, P.O.B 16131
Tel-Aviv, 61161, Israel
rwagner@mta.ac.il

Abstract

We analyse the distribution of distances, in terms of number of turns, between lines in axial maps. We show that this distribution is `scale free', and uniform up to two parameters. We then demonstrate a correlation between the parameters governing this distribution and the parameter governing the polynomial distribution of line lengths in the map. This provides syntactic support to the metrically based categorisation of cities proposed by Carvalho et al. (2003). Finally, we attempt to explain the categorisation in functional terms. To this end we introduce a notion of *attraction cores*, defined in terms of aggregations of random walk agents. We demonstrate that the number of attraction cores in the city correlates well with the parameters characteristic of the distance and line length distributions. Under the conjecture that attraction cores are indeed empirically related to functional centres, we have an interrelated metric-syntactic-functional characterisation of cities.

Introduction

The first goal of this paper is to point out structural regularities in axial maps of cities. We will study the question: How many turns must one take to get from one axial line to another? More precisely, for any integer n we will study the probability that it takes a minimum of exactly n turns (steps on the axial graph) to get from one random axial line to another. In more technical terms, we characterise the *distance distribution* of the axial graph: the probabilities that random nodes in the axial graph are distance n apart for integer n's. This set of probabilities turns out to be highly constrained, and in some sense approximately uniform for all cities considered in this study.

This paper does not only claim that the statistical law of distances in axial graphs is highly constrained, but also provides an explicit model which approximates the distance distribution between random nodes. The model distribution is a *rescaled Poisson* distribution, calibrated by two parameters. It is important to note that this distribution

distribution is not the highly concentrated Gaussian distribution typical of 'small worlds'.

This result alone would have little value, if it weren't related to other features studied by Space Syntax: science cannot be reduced to naming phenomena; at the very least it should also relate them. This paper will therefore demonstrate a link between the above question: How many turns does it take to get from one random axial line to another? And a seemingly unrelated question: How long is a random axial line?

The statistical answer to the latter question (the *line length distribution*) is already known to be highly constrained. In Carvalho et al. (2003) this distribution has been shown to be well approximated by a polynomial distribution. The present paper shows that the responses to the two questions above are correlated. In qualitative lay terms the statement is as follows: the harder it is to find a significant variety of line lengths in the map, the easier it is to observe a significant variety of graph-distances between pairs of nodes in the axial graph. In more precise terms, the more concentrated the (polynomial) distribution of line lengths for a given city, the less concentrated is the (rescaled Poisson) distribution of graph-distances between the city's lines. In fact, we find a significant correlation between the degree of the polynomial which characterises the distribution of line lengths and the parameters which determine the concentration of the rescaled Poisson distribution of distances.

Finally, and most importantly, we attempt to articulate a functional, rather than statistical, correlate of the emergent categorisation of cities. Our conjecture is that cities, where graph distances between random nodes are more tightly concentrated around a typical value (or, equivalently, the length distribution is more spread out), are dominated by a strong global centre. On the other hand, where graph distances are more spread out (and the line lengths distribution is more concentrated) the city will be characterised by a graded hierarchy of local-to-global centres. These phenomena may be related to cultural, historical, topographical or planning-related circumstances.

To test our conjecture, we perform a *random walk* based experiment on the axial graphs. The findings confirm that in the former category of cities the random walk tends to converge towards few global centres, whereas in the latter category the random walk tends to disperse among many local centres. While the experiment is not a proof of the conjecture (such a proof must consider actual cities, and not just axial maps), it does provide it with significant support.

The Model and the Quality of Approximation

The object we wish to approximate is the distribution of graph distances between random nodes in axial graphs (or, equivalently, the least number of turns along paths from one random line to another). To compute this we computed first the distances between all pairs of nodes in the graph. Then, for each node we listed how many nodes lie at each distance from that node, obtaining a list of integer values. Finally, we took the average of all these lists, and divided by the total number of axial lines in order to get a sequence of values which sum to 1. This normalised average will be referred to as the *observed distribution* of distances between random nodes in the axial graph.

To understand the observed distance distribution, consider Figure 1 above (left hand graph). The value at n=4 is almost 0.07. This means that the probability that one would require a minimum of exactly 4 turns to get from one random axial line to another in Las Vegas is almost 7%.

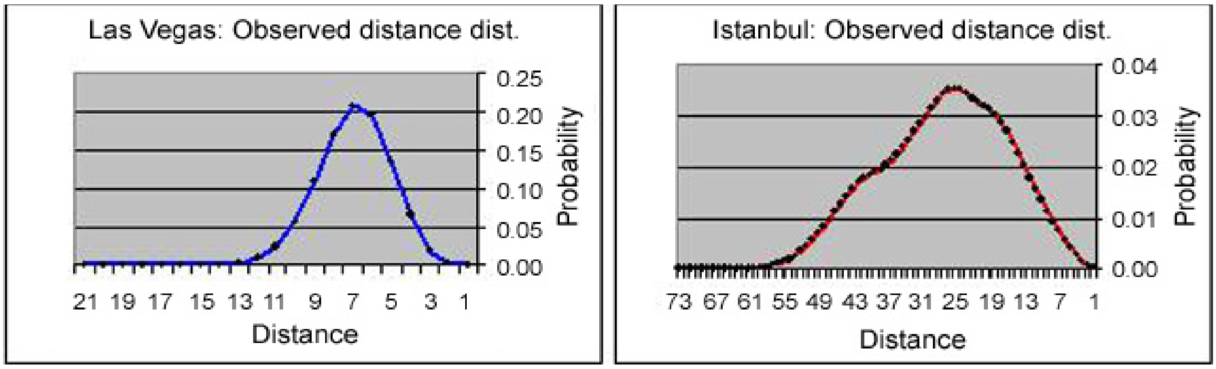


Figure 1:

Observed distance distributions for the Las Vegas and Istanbul axial graphs. The hump on the down-slope for Istanbul appears to be due to the Bosphorus splitting the city

The model which we consider is the following rescaled Poisson density function: $P_\lambda(n) = e^{-\lambda}\ \lambda^{n/s}/s\Gamma(n/s+1)$, where λ and s are parameters, and the Γ function is the standard extension of the factorial to non-integer values, where $\Gamma\ (n+1)=n!$ for integer n (actually, this rescaled Poisson is not exactly a distribution, because the values do not sum to exactly 1; the error, however, is negligible compared to the error in approximating the observed distributions, so I do not bother renormalising).

$P_\lambda(n)$ will be shown to approximate the probability that it would take a minimum of exactly n graph steps to get from one random line in the map to another. The value of s measures the level of *concentration* of the distribution compared to a standard Poisson distribution. For a given value of λ, a lower value of s means a more concentrated distribution (the density function appears more `squeezed'). In other words, fewer distances (*n*'s) have a substantial probability of occurring empirically. A higher value of s means a less concentrated distribution (the density function appears more `stretched-out'), or equivalently, more distances can feasibly occur empirically. λ determines the *shape* of the density function curve. To understand its role better, we note that increasing λ reduces the distribution's concentration, and that the product λs approximates the mean graph-distance between random axial lines.

Table 1 below lists the 30 sampled cities, their sizes in terms of axial lines, the optimal λ and s parameters for each city and indicators for the quality of approximation of the observed distance distribution by the model distributions.

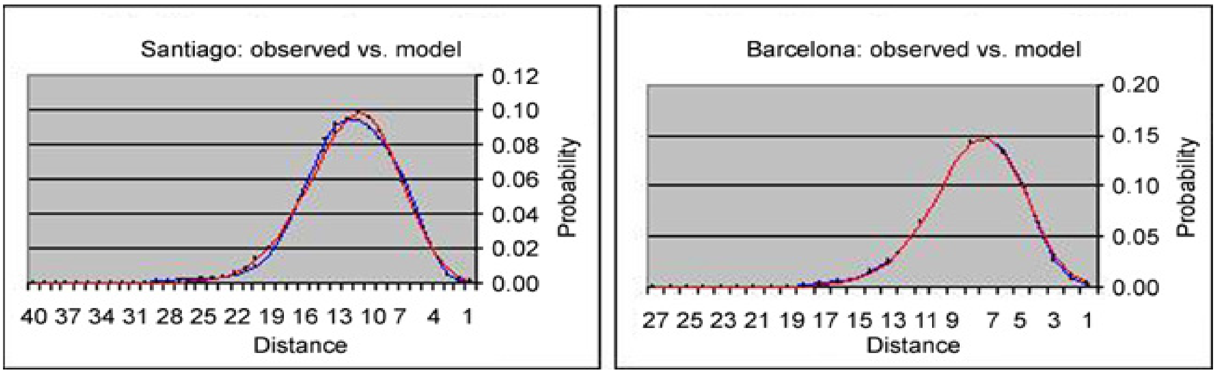
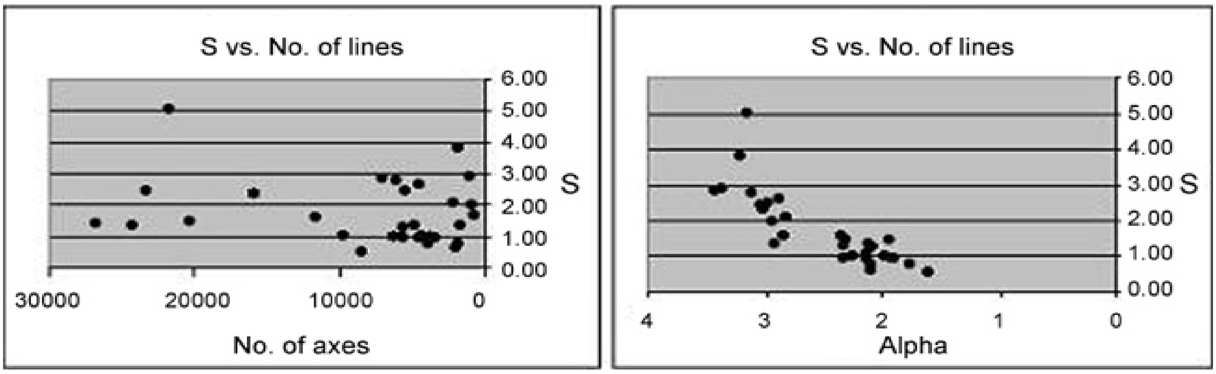


Figure 2:

Caption: Cases of a relatively good fit (Barcelona) and a relatively poorer fit (santiago) against rescaled Poisson (observed distributions in blue, model distributions in red)]

Correlations between the Distance Distribution Parameters and a Metric Typology of Cities

First we must note that the analysis we are doing has nothing to do with the number of axial lines in the maps. The graph in Figure 3 demonstrates the lack of correlation between the number of axial lines in each map and the scaling parameter s (r^2=0.05; if we compare the number of axial lines to λ we get r^2=0.002). What we are measuring here is therefore a phenomenon which does not depend on the sizes of cities, which range in our sample from 854 to 26,281 axial lines.



008-04 ***Figure 3:***

A plot of the parameter s against the number of axial lines in each city (left, r2=0.05) and against the α parameter (right, r2=0.69)]

We also have negligible correlation between the number of lines and the quality of the fit. The correlation between number of lines and the average absolute error percentage from Table 1 is r^2*=0.096*. The quality of the fit does not significantly improve or deteriorate with the number of lines in the range covered by the database.

City	No. of axial lines	s	λ	r^2 of Observed distribution vs. Rescaled Poisson	Mean absolute error percentage of Observed distribution vs. Rescaled Poisson	α	No. of cores
Las Vegas	8370	0.52	13.46	1.000	1%	1.60	1
Semnan	1770	0.78	9.00	1.000	2%	1.75	1
Barcelona	5575	0.95	8.18	0.999	3%	1.91	1
Kerman	4371	0.91	8.80	0.997	8%	1.92	1
Seattle	20213	1.48	9.56	0.997	14%	1.94	5
Qavzin	3723	1.00	9.33	1.000	2%	1.99	1
Milton Keynes	5581	1.28	10.25	1.000	3%	2.08	8
Kermanshah	1870	0.59	12.33	0.999	5%	2.09	1
Hamedan	3855	0.77	10.73	1.000	3%	2.10	3
Bangkok	24219	1.37	9.06	1.000	5%	2.12	6
Shiraz	6258	0.97	9.55	0.985	7%	2.12	1
Pensacola	4296	1.05	8.84	1.000	2%	2.13	2
Amsterdam	9776	1.01	10.85	0.994	9%	2.24	4
Santiago	26821	1.44	8.09	0.989	12%	2.31	4
New Orleans	4846	1.32	6.17	0.925	19%	2.33	1
Hague	3350	0.95	8.74	0.999	3%	2.34	2
Baltimore	11636	1.59	7.35	0.998	5%	2.74	4
Norwich	2119	2.09	6.03	0.999	3%	2.83	7
Winchester	616	1.62	5.76	0.998	3%	2.84	5
Nottingham	4365	2.61	5.30	0.995	9%	2.88	5
Dhaka	1566	1.33	7.54	0.998	4%	2.93	4
Hereford	854	1.96	5.92	0.992	4%	2.93	4
Wolverhampton	5423	2.46	6.41	0.981	9%	2.97	24
London	15912	2.33	6.70	0.999	5%	3.01	13
Athens	23328	2.44	6.39	0.995	3%	3.04	12
Nicosia	6023	2.80	6.37	0.998	5%	3.11	17
Istanbul	21798	5.02	5.35	0.989	11%	3.15	52
York	1773	3.82	4.16	0.840	11%	3.21	10
Hong Kong	916	2.90	3.96	0.984	5%	3.37	2
Bristol	7028	2.85	6.76	0.991	5%	3.42	18

The second column of this table presents the number of axial lines in the axial map of each city.
The third and fourth columns represent the values of s and λ which yield the optimal fit between the observed distribution and the model rescaled Poisson distribution (least absolute residuals fit excluding distances (n's) where the observed value is smaller than for n=1).
The fifth column lists r^2-value of the approximation of the observed distribution by the rescaled Poisson model (again, disregarding distances (n's) for which the observed distribution values were smaller than for n=1 were omitted).
The r^2 values may be misleading here, since the approximation we investigate has little to do with the statistical situation for which the r^2 measurement was designed. The sixth column therefore includes a more `honest' estimate of the fit: the average absolute error in percentage between the observed distribution and the model. Note that readjusting the values of λ and s so as to optimise this measurement of approximation may slightly improve these figures. Here I omit n's for which the observed value was smaller than for n=3, as these values are rarely modelled well.
The seventh column reports the α parameter from Carvalho et al. (2003). The eighth column reports the number of attraction cores defined in the final section of this paper.

Table 1:

List of sampled cities

Recall that the line length distribution was shown in Carvalho et al. (2003) to be approximately polynomial, with a distribution density function of the form $p(x)=cx^{-1-\alpha}$ for some value α characteristic of each city. The most important positive observation here is that there is a significant correlation (r^2*=0.69*) between each of the parameters of the best fitting rescaled Poisson (λ and s) and the α parameter which characterises the distribution of line lengths. Values of α for cities in the sample are included in Table 1.

To appreciate this result, it is crucial to understand that we are dealing with two extremely different parameters: one derives from the metric structure of cities, while the other derives from the syntactic structure of the axial graph, which does not *explicitly* include any geometric information. What is shown here, as well as in other space syntax studies, is that the axial graph *implicitly* codes geometric information, mediated through the structural constraints of actual cities.

Note that the two most obvious outliers for the s against α fit (Figure 3) are Istanbul and York, which also appear to have the most distorted observed distance distribution graphs. Note also that three of the worst fits (Istanbul, New Orleans and Seattle) are split by a body of water with relatively few bridges over it. It might be interesting to attempt accounting for the atypical structure presented by such cities.

As a technical aside, I would like to point out that for some cities the Weibull distribution improves the fit compared to the rescaled Poisson, but that the overall performance of the Weibull distribution is worse than that of the rescaled Poisson. On average, where the Weibull model improves the fit, the improvement is small, while where the Weibull model reduces fit quality, the `damage' is more significant. The Weibull distribution usually turns out to be a better model in those cities which Carvalho et al. (2003) referred to as `red' (α~3). This might further substantiate that paper's blue/red categorisation of cities into two distinct growth models.

Can the Statistical Metric-Syntactic Analysis Tell us Something about the Functional Structure of the City?

The first thing to note, on the functional level, is that cities do not conform to standard *Small World* models. Typical Small World graphs show an approximate Gaussian distance distribution concentrated tightly around very few values (for this fact the reader may consult Krioukov et al. (2001), as well as Dorogovtsev et al. (2002) and Krzywicki (2001)). The fact that cities are not small worlds is also reflected by the fact that the distribution of degrees *connectivity*, in space syntax jargon) in axial graphs is not power-law (according to my own tentative calculations, it can be approximated by a power-law/exponential mixture, but the results are not very impressive).

The fact that cities are not small worlds corresponds to a basic tenet of Space Syntax: it is because cities have a wide variety of differently positioned spaces that they manage to act as socially productive machines. A small world, where almost any position is almost equally distant from almost any other position, would not be up for the task.

On the other hand, the distance distribution of axial graphs is `scale free' in the sense that the number of axial lines does not significantly affect the shape of the distribution. These facts suggest that if we wish to model axial maps, we should probably look away from standard small world models and focus more, perhaps, on copying models. Unfortunately, there is not a lot of relevant knowledge concerning copying models in the literature. The existence of a categorisation, which is reflected by both metric and syntactic statistical data, suggests a more tangible underlying feature. In Carvalho et al. (2003) the authors observe that cities with lower α have more long integrators crossing through them.

To elaborate on this conjecture; the first category (higher values of s and α) consists of cities which appear to be more spread out, and broken into a larger number of functional sub-units. This may correspond to a city formed by connecting several once-independent communities (if we allow schematic oversimplification, London can be taken as an example). The cities in the second category (lower values of s and α) seem to function more homogeneously around a strong unified core. Las-Vegas could serve here as an example. In other words, the conjecture is that cities in the first category have a distributed structure, with a hierarchy of local sub-structures, whereas cities in the second category have a more centralised structure.

The line of thought in this study, which at this point is still tentative, is as follows. Imagine first a city without significant local centres. Randomly chosen axial lines are very likely to be peripheral lines (because there are much more peripheral lines than central ones). The shortest path between them is likely to go through the centre, and the distance between them would be close to the sum of their distances to the centre. This dictates a relatively concentrated distance distribution. In a city with a graded hierarchy of centres, on the other hand, an optimal trip from one peripheral node to another has a non-negligible likelihood of passing only through local centres, rather than through the main global centre. This contributes towards a greater variety of distances between nodes.

This understanding is related also to line lengths. In a city with a single dominant centre, we will have relatively more long axial lines connecting the periphery to the dominant centre. In a city with a graded hierarchy of centres, however, the connection between periphery and centre will be broken into smaller segments advancing via local sub-centres. The result would be a significantly smaller proportion of long lines. This structural difference could be related to historical, topographical, cultural and planning related circumstances. Investigating such circumstances is beyond the scope of this paper. I would like, however, to describe an experiment which appears to supports my conjecture.

The experiment is based on the concept of a *random walk*. Consider an agent standing on a node of a graph. Let the agent choose with equal probability any of the edges connected to his node, and then traverse the chosen edge to an adjacent node. It is a well known and fundamental fact (the so called *ergodic theorem for reversible Markov chains*) that as the number of steps increases, the portion of steps spent at each node converges toward a number proportional to the degree (or, in space syntax jargon, connectivity) of that node, provided, of course, the graph is connected.

The experiment I suggest is a variation on this theme, but can also be thought of as an iterated version of the *control* measure from Hanson and Hillier (1984). Suppose that on each node one has a large number of agents. Suppose that at each step the agents proceed to adjacent nodes, but that their choice of target nodes is not uniform; instead, the distribution of the agents' target nodes is proportional to the actual number of agents on each of those target nodes. For example, suppose that a given node has 100 agents in it, and that it has three adjacent nodes containing 90, 150, and 60 agents respectively. Then in the next step the 100 original agents will be distributed between the adjacent nodes proportionally to the ratio 90:150:60. More explicitly, 30, 50 and 20 agents will advance from the original node to the first, second and third adjacent nodes respectively.

This process simulates a random walk with a multiplier factor. The more occupied a node is, the more likely agents are to go there. After many steps, most agents end up congregated in a few *attraction cores*. The London axial map in figure 4 shows the most occupied lines containing altogether 50% of the original agents after many steps of our walk (the walk is started with an equal number of agents on each line, and repeated for 300 steps, at which point the distribution no longer changes significantly. The choice of 50% is arbitrary, and significant results can be obtained with other thresholds as well). In London these lines form 12 connected attraction cores containing 46 axial lines. The results for other cities are presented in Table 1.

Figure 4:

The attraction cores of London

What is the meaning of such an experiment? At this point it is not completely clear. Obviously, the random walk described above cannot presume to reflect exactly the actual movement of human agents. Moreover, its sensitivity to the initial distribution of agents and to errors in the production of the axial map is yet to be tested. Some preliminary tests indicate that the results are reasonably stable, but these tests are far from a systematic study[i]. More importantly, the correspondence between the attraction cores picked up by the experiments and actual functional phenomena has not been addressed at all beyond a preliminary gaze[ii].

The number of attraction cores, presented in Table 1, correlates strongly with the other parameters considered in this paper, and do not depend heavily on the size of the city. Indeed, there is only a mild correlation between the number of lines in the city and the number of attraction cores (*r^2=0.37*; some correlation is expected, because it is more difficult to maintain a unified core in larger cities), whereas there is a strong correlation between the values of the parameter s and the number of attraction cores (*r^2=0.80*). This means that there is a strong quantitative connection between the number of cores and the distribution of distances between nodes in the axial graph. It would, however, be premature to state that we have positively related our syntactic-metric structural analysis of cities to an actual functional feature. To demonstrate such a relation one must look beyond the axial map.

This paper, then, leaves several interesting questions open. What are the functional significance of attraction cores as defined in this paper?

How stable is their detection? Can the syntactic-metric characterisation of cities advanced in Carvalho et al. (2003) and in this paper be related to functional features of cities or to their historical development? And a further question, as suggested by Alan Penn: can we relate our findings to findings on functional hierarchies stemming from studies such as Berry and Garrison's (1958)?

Acknowledgment; The computations for this paper were performed using the application DataMAPS written by the author's M.Sc. student Gal Briner and MATLAB 7. I would like to thank Gal Briner for his hard work. I would also like to thank Alan Penn, Rui Carvalho, Shinichi Iida and Bill Hillier for useful comments and access to datasets.

References

Berry, B., Garrison, G.L., 1958, “The Functional Bases of the Central Place Hierarchy”, *Economic Geography*, 34, 145--54

Carvalho, R., Iida, S., Penn, A., 2003, “Scaling and Universality in the Micro-Structure of Urban Space”, J. Hanson(Ed.), *Proceedings*, 4th International Space Syntax. Symposium, University College London, Vol I, paper no. 34.

Dorogovtsev, S.N., Goltsev, A.V., Mendes, J.E.F., 2002, “Pseudofractal Scale-free Web”, *Physical Review E*, 65(06).

Hillier, B., Hanson, J, 1984, *The Social Logic of Space*, Cambridge University Press, Cambridge.

Hillier, B., Penn, A., Turner, A., 2005, “An Algorithmic Definition of the Axial Map”, *Environment and planning B*, 32, pp. 425-444.

Krzywicki, A., 2001, “Defining Statistical Ensembles of Random Networks”, *Workshop on Discrete Random Geometries and Quantum Gravity*, Utrecht.

Krioukov, D., Fall, K., Yang, X., 2004, “Compact Routing on Internet-Like Graphs”, in Proc. IEEE INFOCOM.

Peponis J., Wineman J., Bafna S., Rashid M., Kim S.H., 1998, “On the Generation of Linear Representations of Spatial Configuration”, *Environment and planning B,*25, pp. 559-576.

1. (The problem of the sensitivity of space syntax measures to the choices and errors of the person drawing the axial map is more general. A significant move towards a solution was made with the introduction of algorithms for the production of axial maps (Peponis et al., 1998 and Hillier et al., 2005). Nevertheless, most axial maps in existence today are `hand drawn'. Therefore, measures which are relatively new, such as line length distribution, distance distribution and attraction cores, should not be uncritically assumed to be insensitive to the choices and glitches of the person drawing the map.)

2. It seems that attraction cores capture important junctions (rather than axial lines), and that these junctions indicate the existence of a distinct spatial unit around them, without necessarily capturing all functionally significant junctions.