To tame a TIGER one has to know its nature[1]:
Extending weighted angular integration analysis to the description of GIS road-centerline data for large scale urban analysis

65

Nicholas S. Dalton[1], John Peponis[2], Ruth Dalton[2]
[1]Georgia Regional Transportation Authority, USA
[2]Georgia Institute of Technology, USA

Keywords
Axial maps; GIS; TIGER files[2]; fractional depth, angular analysis

ndalton@grta.org
john.peponis@arch.gatech.edu
ruth.dalton@arch.gatech.edu

Abstract
GIS databases representing urban layouts according to road centerlines spanning between intersection nodes (at road junctions) can be analyzed syntactically based on the concept of angular fractional depth.

Introduction

The automation of axial representations of urban systems has been a desideratum in the field of space syntax for at least two pragmatic reasons, over and above the theoretical advantages regarding mathematic definition and formal elaboration that would be implicit in a fully algorithmic procedure of general applicability. First, the production of axial maps can be a tedious and protracted process. An automatic system would greatly reduce the time taken to bring an area under analysis. This would be particularly significant in the analysis of large metropolitan scale maps, which are not currently amenable to analysis within reasonable constraints of available labor time. If automatic processing was possible, large samples of different urban forms could be efficiently compared, new heuristic models could be developed and research hypotheses could be brought to test more easily. Second, the production of axial maps often involves the judgment of the researcher in the translation of the underlying map to an axial representation. For example, in some cases error may lead to not recognizing that a single line can traverse a given irregular and fragmented area, while in other cases judgment may be needed in order to determine whether a given intersection in the original map should be interpreted as an overpass, a cross-junction, or simply a traversing line with two adjoining dead ends. Thus, situations are possible where two different axial maps might be produced from the same original map. Generally the robustness of integration analysis absorbs any minor differences between axial maps. However, the automatic generation of axial maps would eliminate uncertainties arising from the possibility of random errors or divergent judgment. As Batty asserts in his paper, (Batty, 2000) *"We believe that by introducing* *automatic methods, space syntax will be given a chance to relate to mainstream ideas in morphology and social physics, thus widening its appeal to disciplines beyond architecture."*

This paper discusses a particular approach to automation, which is based on the analysis of GIS databases on major metropolitan areas in the United States. GIS databases often provide large amounts of information on land uses, traffic flows, or the spatial distributions of social or economic phenomena. While the data is not always presented at levels of resolution that would correspond to the spatial elements that are normally at the foundations of syntactic analysis (for example block-level or census track level data cannot easily be associated with properties of the axial map except at a level of aggregation that eliminates some of the discriminating powers that make space syntax an attractive method of analysis in the first place), the advantages of facilitating the potential interface between space syntax and GIS databases are obvious.

The main problem associated with the application of syntactic analysis to the representation of streets in most GIS systems arises from the fact that the latter are based on road centerlines extending between intersection nodes. Consistent with this, the long linear stretches of road that are so frequently found in cities in the US appear segmented. Consequently, their analysis by the traditional methods associated with space syntax makes them appear deep. The integration cores established on the basis of such analysis are highly clustered near the center of gravity of the area and not picking up the global spatial structure. Figure 1, for example juxtaposes an analysis of Downtown Atlanta based on a traditional axial map drawn according to a detailed map of urban layout (Figure 1a) to an analysis of the same area according to centerlines linking junction nodes in a GIS database (Figure 1b). The first analysis reveals an integration core, which includes traversing as well as peripheral streets and linking the various parts to each other. The second analysis reveals a structure of concentric zones of spaces with similar integration values, with more integrated segments near the geometric center of the area and less integrated segments at the periphery. This shows graphically that traditional syntactic analysis applied directly to TIGER representations of US cities will give results that are quite misleading as compared to the results that would have been obtained by normal syntactic analysis. Additional discussion of this is offered below.

However, the mode of representation that is typical in GIS databases also has some advantages. For example, it lends itself to discriminating conditions along different portions of a long road (for example changes in land use, or changes in the frequency of intersections, or, indeed, changes in vehicular or pedestrian traffic densities). If different centerline segments were merely joined into a single axial line based on normal conventions, the ability to represent such changes of condition would be lost. Thus, there are incentives to seeking ways to apply syntactic analysis to road-centerlines extending between intersection nodes, without concatenating them into axial lines as normally defined in space syntax.

An additional incentive for dealing with road centerlines is that they are much more broadly available than building outlines, property outlines or block outlines. If there were ways to deal with them sensibly within the framework of space syntax, analysis could be applied to systems for which the information normally taken into account in order to derive the axial map is missing.

Consistent with the above, we discuss a way in which data available in TIGER format can be analyzed syntactically. The TIGER format is typically used not only in the United States but also internationally.

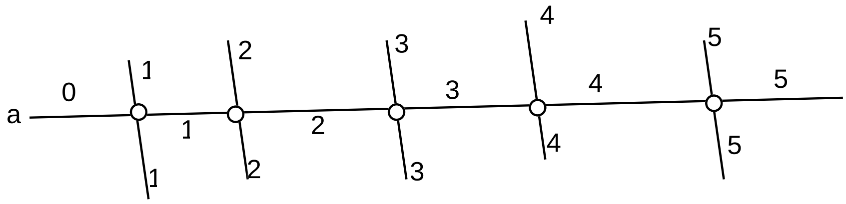
a b

Figure 1:Downtown Atlanta represented as an axial system (1a) and as TIGER centerlines (1b), both analyzed according to standard syntactic techniques.

The key to analyzing TIGER files: fractional depth

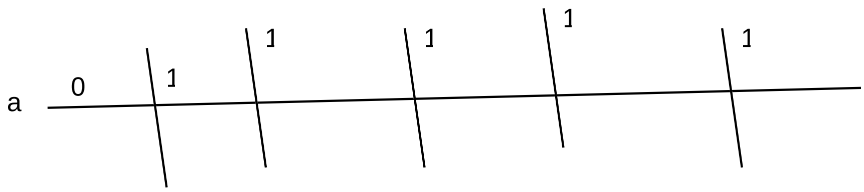
Figure 2a represents a street as a series of line segments joined at junctions. The street end on the left (point a) is treated as a starting point to assign depth values to all other segments according to standard syntactic analysis: a unit of depth is added at every transition from one segment to each of the segments accessible from it. In Figure 2b, the same street is represented as a single line rather than six line segments.

The intersecting streets are also represented as single lines rather than as two line segments each. Depth values are assigned from the point of view of the same street end. The difference between the depths produced by the two representations summarizes the problem that needs to be resolved in order to render TIGER files amenable to syntactic analysis. A TIGER map would represent this condition according to Figure 2a rather than 2b, and therefore lead to exaggerated depth values as well as distorted distributions of integration, as shown in Figure 1.



2a

Figure 2: A long street intersecting a series of short streets represented segmentally (2a) and axially (2b) and analyzed according to standard syntactic techniques, showing depth gain from a starting point, a.



2b

The augmentation of depth that would result from the application of normal syntactic analysis to TIGER files is summarized in Table 1, with respect to the downtown area of Atlanta shown in Figure 1. The table clearly shows that the numbers of segments used to cover the area differ by orders of magnitude, and that the integration of the area is much higher under normal syntactic representation than it is under TIGER representation. In addition, however, the table shows a significant drop in the correlation between connectivity and integration when we shift from traditional to TIGER representation. This is probably associated with the fact that the number of intersections associated with each segment varies more in the traditional representation than in the TIGER representation.

Dalton (2001) has proposed that depth can be measured according to fractional rather than unit changes. In standard syntactic analysis, the transition from one line to another is treated as a change of direction and is registered as a full depth value. Dalton proposed that the transition should be treated as a fraction between 1 and 0 depending on the angle at which the two lines intersect. Where the lines intersect at right angles the fractional depth gain should be 1 and where they intersect at 180^{o} it should be 0.

System	# Spaces	Average Mean Depth	Average Connectivity	Inverse Correlation between Mean Depth and Connectivity/r
Downtown Atlanta Axial	1005	7.513	4.137	0.457
Downtown Atlanta TIGER (Traditional)	3030	27.339	4.669	0.247

Table 1: A comparison between an axial map and a TIGER map analyzed according to standard syntactic techniques.

The application of fractional depth opens the way for the analysis of TIGER representations of urban layouts. Where line segments are almost co-linear, depth gain under fractional analysis would tend towards 0. In effect, this means that almost collinear line segments are effectively treated as the single line of the traditional axial map. This is clearly showed in figure 3 which attributes fractional depth to the elementary pattern previously discussed in Figure 2. When we take the line end on the left (point a) as a starting point the depth values associated with each line segment are almost identical to the values that would have been associated with the corresponding axial line under the traditional representation.

We have, therefore, sought to explore whether the idea fractional angular depth, as proposed by Dalton (2001) opens the way for the syntactic analysis of large systems represented in TIGER format.

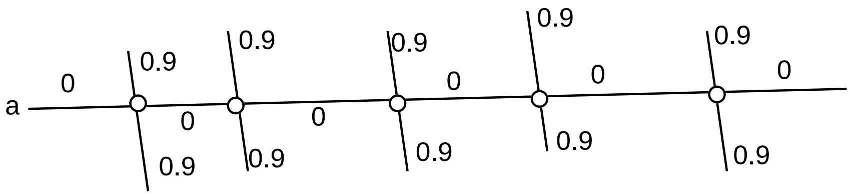


Figure 3: The line system from figure 2 analyzed fractionally and showing fractional depth gain from a starting point, a.

Some technical considerations

To test this concept, a new software application, *Tigerman*, was written. The data is first exported from the GIS system as a series of line segments. Each line segment maintains an id number from the original GIS system. This value is preserved to help match the numerical output of syntactic analysis back into the GIS data later on.

The TIGER format road data is characterized by some interesting advantages and anomalies that must be dealt with. The main advantage, at least in principle, arises from the manner in which nodes are defined. By normal TIGER conventions, every time two roads cross, their intersection is represented as a node linked to four independent line segments. Thus, if two lines cross at different grades, as with an overpass, no node will be shown. By getting *Tigerman* to only recognize intersections where nodes exist, the problem of dealing with overpasses is automatically dealt with. When lines appear to intersect without having a node represent their intersection, the intersection is treated as an overpass/underpass/bridge and thus not taken into account in the calculation of syntactic depth. No special action, such as unlinking lines, is needed. In this respect, *Tigerman* is more efficient than other software applications performing traditional axial analysis such as *Axman, Spatialist, MeanDA or WebMap*.

The main disadvantage of analyzing TIGER data arises from flaws in the way in which data is input. Specifically, the endpoints of two segments that meet at node might not be coincident but might be offset by very small amounts, due to errors in original data input. That is to say, one line segment might terminate at coordinates 10222.0222, 3329983.2 and another might begin at coordinates 10222.0221, 3329983.2. This could lead to a failure by the algorithm to recognize this intersection when depth is calculated. To deal with this, *Tigerman* includes a

module which searches in the vicinity of the endpoints of a line segment looking for possible endpoints of other line segments which are sufficiently close to suggest a poorly represented intersection. When it finds them, it adjusts the data to recognize and include that intersection. This module works well in ensuring that the continuity of the urban fabric is properly maintained where it should be. However, there is a risk that connections are made in the data where they do not actually exist. However, the risk of "over-connecting" the system is reduced when we bear in mind that the module looks for proximity between endpoints only rather than more generic proximity between endpoints and line segments. At present, *Tigerman* has a particular metric radius value (e.g. 3 meters), which it uses to search for these 'near' misses. This value can be adjusted empirically: if too small, then it will exclude legitimate connections; if too high it will result in finding false connections. This kind of problem is common when maps are used as part of urban transportation models. In the long term, the problem may be eliminated by very carefully editing the map. At present, however, the risk that *Tigerman* might lead to misrepresenting the system in particular situations is accepted as a legitimate source of potential error; it is anticipated that researchers will attempt to limit this type of error so as not to affect the overall reliability of the analysis.

A third problem concerns interpretation rather than data input. TIGER data describes road centerlines. *Tigerman* assumes than all these lines are valid routes for movement. This means that if *Tigerman* is used to analyze an urban system in order to correlate spatial variables with data on pedestrian or vehicular movement, some of the lines included in the analysis will not carry pedestrian or vehicular loads in reality, because they may correspond to freeways, access ramps, or conversely pedestrianized streets. Again, this is an acceptable problem that could only be controlled where all line segments are labeled or weighted according to the kind of movement that is associated with them.

Tigerman runs under the Linux/Unix operating system. Given the change from integer to floating point arithmetic and the processing of large numbers of segments, the processing is much slower than for traditional syntactic analysis. Current computing hardware is now making possible the kind of computation required by *Tigerman*. During testing it was necessary to develop a number of algorithmic improvements that reduce the computation time. If processed in a manner identical to Axman, systems such as the Metropolitan Atlanta Area would take several months or even years to complete. After logarithmic improvements, a PC running with 64 MB of RAM takes 7-14 days to process a large-scale metropolitan area consisting of 600,000 segments. Faster computing might reduce the time taken to process a large metropolitan area down to a single day. This might seem like a long time but it must be remembered that the information does not have to be manually digitized.

Pilot analyses of Atlanta – the problems of defining radius and relativization functions

The first results of processing TIGER files are presented in Figure 4. The system under analysis is once again Downtown Atlanta and hence is intended to be compared directly with the two illustrations in figure 1. By inspecting Figure 4, we can easily see that the integration analysis, at infinite radius, has picked out several significant arterials in the urban system. The pattern of integration is far more similar to that of Figure 1a (integration analysis of an axial map of Downtown Atlanta) than that of Figure 1b (traditional analysis of an TIGER map of Downtown Atlanta).

It is also possible to reproduce Table 1, but this time to add an additional row illustrating the analogous results produced by a *fractional* analysis of the TIGER data. This new data is presented in Table 2. Now, it can be determined that although there are more than three times as many lines in the system (since each original axial line may be represented by numerous TIGER centerlines), there is not the equivalent overall gain in depth that occurred when the TIGER data was processed using traditional space syntax methods. In fact, the average Mean Depth of the system, processed fractionally, is more similar to the average Mean Depth of the original axial system than of the TIGER system processed using traditional methods. Equally, the correlation, r, between the reciprocal of Mean Depth and Connectivity is closer to the correlation between traditional axial analysis than to the first TIGER analysis.

Why are we looking at values of Mean Depth rather than integration in Tables 1 and 2? The reason is that although we can compare the changes in integration between an axial representation of Downtown Atlanta processed traditionally and the TIGER data for Downtown Atlanta. We are unable to compare these results with the TIGER data of Downtown Atlanta processed fractionally. This is because the calculation of the measure integration relies upon a process of relativization and the equivalent relativization required to produce a measure of fractional integration is rather less straightforward. The next section of the paper outlines the particular problems associated with radius and relativization for fractional systems.

65.8

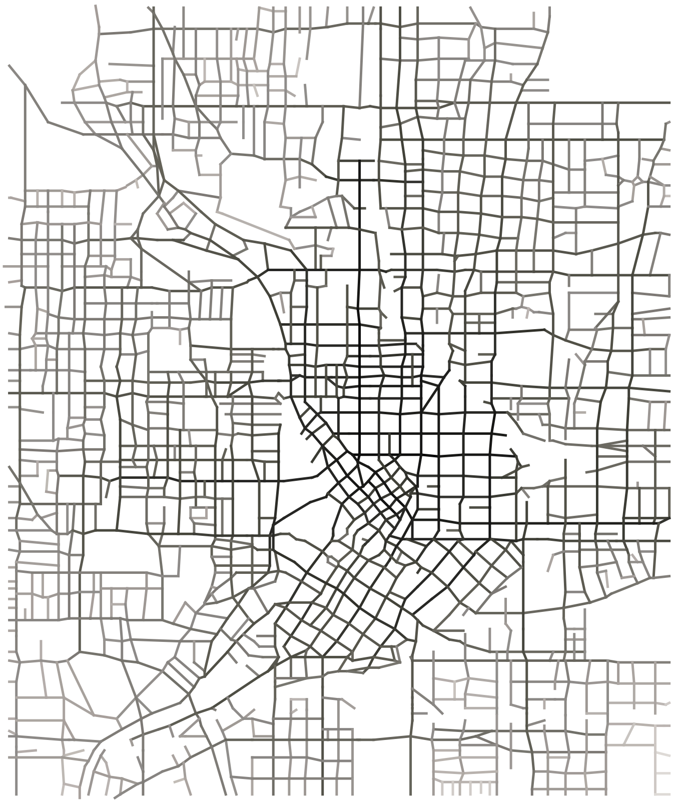


Figure 4: Downtown Atlanta represented as TIGER centerlines analyzed according to new fractional techniques.

As we move into the analysis of very large systems, infinite radius analysis is not likely to be particularly revealing of urban function, unless large scale movement and land uses of metropolitan significance are under investigation. For other purposes, including the analysis of pedestrian flows, it is important that the radius of analysis be specified. In traditional syntactic analysis, radius is defined according to the number of steps that are taken moving out from each of the component lines of the axial maps. Given that *Tigerman* operates with segmented data to perform fractional analysis, new questions arise: how should the idea of radius be re-defined so as to incorporate the concept of angular fractional depth? How can we specify a "movement horizon" taking the fact the representation of the street layout is highly segmented in the first place? We decided to specify radius according to the angular sum of direction changes involved as one moves out from a line segment. Various limits were tried, for example 180^{o} or 270^{o}. In the even, an aggregate angle of 2.27 radians (130°) was fond to "pick up" many of the significant roads in the system, and not only the main arteries. Thus, from an intuitive point of view, 130° seems like a potentially relevant radius threshold, whose significance has to be tested in the future, based on correlations with empirical data on movement, pedestrian or vehicular.

System	# Spaces	Average Mean Depth	Average Connectivity	Inverse Correlation between Mean Depth and Connectivity/r
Downtown Atlanta Axial	1005	7.513	4.137	0.457
Downtown Atlanta TIGER (Traditional)	3030	27.339	4.669	0.247
Downtown Atlanta TIGER (Fractional)	3030	6.417	4.669	0.337

Table 2: A comparison between an axial map and a TIGER maps analyzed using both standard and fractional syntactic techniques.

A more interesting theoretical problem, however, also arises as we deal with the question of radius. The introduction of radius creates the need for a method of relativization. In traditional syntactic analysis, Average Mean Depth is adjusted to take into account the number of elements involved in the calculation and the adjustment is incorporated into the measure of integration. Thus, it is possible to compare the syntactic properties of elements whose integration measure is calculated by taking into account differing numbers of other elements that are accessible within a given radius. Relativization is fundamental to space syntax as an analytic theory precisely because it allows us to compare not only systems of different sizes, taken as whole, but also parts of the same system that involve different numbers of elements. Given this, how should we handle relativization when dealing with highly segmented data and with fractional angular depth? If we consider a line made from a number of segments such as in Figure 3, for example, we could have a situation where the total depth gain as we move along is zero, yet the number of segments is a number, n, greater than 0. This situation would never arise in traditional syntactic analysis.

We have not yet developed appropriate relativization functions for TIGER syntactical analysis. To give an impression of what a radius function might look like it is possible to bypass the relativization problem completely. When dealing with the analysis of a single system at radius infinity, for example, relativization is not needed because the number of elements considered is constant and the relative integration of individual spaces can be characterized according to the corresponding total depth of the system. The value of total depth is then a *pure indicator* of the shape and size of the justified map from that point. In the special case of a large system we can discover that for a given depth, radius R, that every single segment in the GIS map will have at least K lines within that radius. Remember that R is a measure of the cumulative angle in radians not in total steps of depth. For a value of R near 3.14 radians (180°) we can expect to find at least K (for example, 50) spaces accessible from the most segregated space in the system.

We define our mean depth for radius R as the total of the smallest K depth values from a given point, divided by K. For example if K is 50, we examine the bottom 50 smallest depths from a starting point P within radius R. By fixing K to be

50, we sidestep the problems of relativization. However, this is not a general solution if we cannot guarantee that all urban maps will have at least K connected spaces with R steps/radians from any starting node. While this solves the problem of radius for a single system, it does not necessary make a comparative mechanism between cities possible (unless R and K are appropriate for both). Another potential problem arises if K particularly small: in this case the variation of values over the whole system will be too great and hence will not be representative. This is similar to setting the value of R to be too small. By empirically testing the system with various values for R, we can determine the largest permissible values for K. The map can then be reprocessed for that value of R and maximum K. Eventually; if enough example cities with movement data are available we can find a value for R that serves as good predictor of pedestrian movement (analogous to radius 3), cycle movement (analogous to radius 5) and to vehicular movement (analogous to radius 7).

Conclusion

Although the work presented in this paper is still at an early stage, the analyses of Downtown Atlanta and the results presented in Table 2 are extremely encouraging. It does appear that by applying methods of fractional analysis to TIGER data, as supplied by many GIS systems, that a close approximation to traditional axial analysis can be achieved. As discussed at the beginning of this paper the benefits of this analysis are manifold, from the ease of automation, to issues of standardization, pedestrian movement prediction, compatibility with other analyses and appeal to other academic disciplines. However, the second half of this paper concludes with a presentation of the issues that arise from the fractional analysis of TIGER data, with respect to relativization and the concept of step-depth radii. The paper begins to suggest ways in which these potential problems might begin to be addressed, but acknowledges that these are vital issues for future work. In summary, this paper presents the first stepping stone in what is anticipated will prove a major breakthrough in analytic methodology in space syntax research and beyond.

Notes

[1] An ancient Chinese proverb.

[2] TIGER is an acronym for **T**opographically **I**ntegrated **G**eographic **E**ncoding and **R**eferencing.

References

Dalton N, 2001, "Fractional Configuration Analysis and a Solution to the Manhattan Problem", *Proceedings of the Third International Symposium on Space Syntax* , Atlanta, GA

Batty, M., Rana, S., *Reformulating Space Syntax: the Automatic Definition and Generation of Axial Lines and Axial Maps*, Centre for Advanced Spatial Analysis Working Paper Series, Paper 58