#150
EMPLOYING VOLUNTEERED GEOGRAPHIC INFORMATION IN SPACE SYNTAX ANALYSIS

KIMON KRENZ
Space Syntax Laboratory, The Bartlett School of Architecture, UCL
kimon-vincent.krenz.12@ucl.ac.uk

ABSTRACT

The application of volunteered geographic information has rapidly increased over the past years. OpenStreetMap (OSM) forms in this context one of the most ambitious and promising projects, providing consistent global coverage of street network information. With a constantly growing number of participants and the implementation of governmental and proprietary based information is a complete coverage of global street networks within reach. The data allows comparative cross-country analyses and any method developed within its framework are transferable to other cases. This makes OSM a powerful and desirable data source for applied network analyses, such as space syntax. However, OSM data does not come without obstacles. Inconsistent representation of space, topological fragmentation and accuracy are just some of the problems that one faces when employing OSM data. In fact, without prior processing and simplification of the network, results differ significantly between case studies. This paper presents a method for OSM data set simplification as well as the theoretical and analytical reasoning behind it. The simplification is done by a series of ArcGIS workflows and algorithms. The outcome of this process is compared to an angular segment analysis (ASA) of a segment model, an Integrated Transport Network (ITN) Ordnance Survey data model and an OSM street network data model. The results show that a simplified version of OSM data is highly comparable to a segmented axial line representation and that such data sets constitute an appropriate alternative for situations where segment maps are not available, such as complex, large-scale regional models and cross-country comparisons. The simplification workflow is transferable to other cases and data sets and helps overcoming common problems while significantly improving computational time needed in the process.

KEYWORDS

Volunteered Geographic Information, Open Street Map, ArcGIS, Space Syntax, Street Network

1. INTRODUCTION

The aim of this paper is to present a workflow and methodology that allows the use of OpenStreetMap (OSM) data in space syntax angular segment analysis (ASA). The reasoning behind employing such data sets is the increasing scale of analytical investigations in the context of space syntax. This augmentation of scale has become particularly necessary due to the extensive global growth of cities and their urban hinterland into large complex urban regions. These urban structures are simply too vast to be mapped manually or generated by automated algorithms. This has created a situation in which the time and economic feasibility of traditional as well as algorithmically derived axial line maps needs to be revisited. Previous research proposed to make use of governmental so-called road-centre line data as an alternative for a segmented axial line, more commonly referred to as segment maps (SM). However, very little has been said about the disadvantages of such approaches particularly when global comparability is needed, something in which space syntax is believed to be particular strong. OSM road-centre line data, on the other hand, I will argue, forms not only an appropriate alternative basis for models in these situations, but it also allows global comparability as well as being freely accessible on a large scale. Nevertheless, OSM data does not come without disadvantages either. Particularly concerning excessive information in such data sets, which makes a simplification prior to any ASA application necessary, caution needs to be exercised.

This paper consists of three parts; the first revisits the foundation of space syntax axial line models and the sequentially developed analytical method of ASA and its segment map (SM) model. An emphasis is placed on the model underlying the analysis and the difficulties arising in the model generation generally and in large-scale applications particularly. In this light, volunteered geographic information and governmental road-centre line data, such as the British Integrated Transport Network (ITN) are reviewed as alternatives for SM models. Finally, advantages as well as disadvantages of OSM data are discussed and the effect of these on ASA outcomes.

The second part presents the structure and particularities of the previously introduced OSM data, as well as the difficulties researchers are facing when employing such data in ASA. I discuss the three main difficulties, which are topological inconsistency, traffic management components and excessive or redundant nodal information. I propose different GIS strategies to simplify and remove such redundant information and explain the theoretical reasoning behind them. The result is a newly derived simplified OSM network model, termed 'SIMP'.

The third part evaluates the new SIMP model against OSM, ITN and SM models in ASA. I do this, using descriptive statistics, visual comparisons, as well as a Pearson and Spearman correlation analysis. The results show an overall high correlation between the four models, confirming previous findings. The new SIMP model exhibits higher correlations with the SAL model than both OSM and ITN network models, indicating that a simplified OSM network does not only form an appropriate alternative but one that presumably incorporates fundamental network characteristics of SM models.

2. AXIAL MODELS AND ANGULAR SEGMENT ANALYSIS

Axial analysis forms one of the fundamental techniques of space syntax. At the core of an axial analysis methodology lies the axial line map, a representation of the continuous structure of open spaces in urban settings. The first axial line model was introduced by Hillier and Hanson (1984, p. 17) during the early 1980's and defined as a system of fewest and longest intersecting lines covering all open spaces. These lines are the result of a two-step process where the spatial system under investigation is first represented through a two-dimensional organisation of convex spaces. Convex spaces are polygonal representations of continuous open spaces, in which each part of a space must be visible from every other part. The underlying rule for drawing a convex space is that each polygon must feature the best 'area-perimeter ratio', starting with the 'fattest'. In a subsequent second step, this system of convex spaces is covered by a one-dimensional set of axial lines. Axial lines are linear representations of longest lines of sight and/or movement. Each convex space must be covered by at least one axial line, while each line needs to be the 'longest straight' line possible (ibid., p.17).

Although Hiller and Hanson describe this process as reproducible and objective, there is some discussion and ambiguity about the comparability and making of axial maps. Problems arise for instance with differences in the level of detail or resolution in which convex spaces are produced, as this impacts the number and distribution of the resulting axial line map. Problems also arise with the difficulty to arrive at comparable reproducible solution for the same given urban context. Peponis *et al.* acknowledge in this regard 'SpaceBox'[1], a software that automated the generative process of convex spaces, but they criticise the mathematical

1 SpaceBox is a software developed by Sheep Dalton (1988) and includes several space syntax related functionalities one of which being the generation of an all convex space map. The software's partitioning algorithm extends a walls surface area collinear until the produced line reaches another wall surface. See Carranza and Koch for more recent work on convex spaces (2013).

rigour of its computational algorithms to generate convex spaces (1997, 1998). According to Peponis et al. neither the initial principle of generating convex spaces based only on an economic partitioning, nor the extension of surfaces to the next opposite wall is a sufficient method. Both lead to multiple, conflicting solutions, implying that a more sophisticated set of rules is necessary. Interestingly, although the methodology of convex spaces is thought of in an urban context most of the discussions are set in the context of buildings. This might be due to the time-consuming process of producing convex spaces for entire cities, with the sole purpose of deriving an axial line map. The scale of the area under investigation and respectively the time necessary to produce such convex representations is certainly one of the most important influencing factors.

Moreover, Desyllas and Elspeth argue that not only the production of convex spaces, in general, is difficult, but that it constitutes a 'mathematically impossible problem' to link all maximal convex spaces with axial lines in an identically repeatable manner (2001, p. 27.6). The core problem here is that there are several solutions to axial lines that fulfil the criteria of being the longest as well as covering all convex spaces (Batty and Rana, 2004; Ratti, 2004). As a solution to this technical and theoretical problem Turner et al. (2005) – building on an initial but not ideal solution from Peponis et al. (1998) – proposed an automated methodology that produces a fewest line axial map. The starting point of their method is vector information of open space boundary polygons. Based on this, a so-called 'all-line map' is generated (Penn et al., 1997). The 'all-line map' is a map that features all lines that connect each vertex of boundaries and buildings with all other visible vertices, i.e. all possible lines of movement. In a following step Turner et al. employ an algorithm to reduce this 'all-line map' to a fewest line axial map. Their results are reproducible and strikingly similar to the original Hillier and Hanson axial map (2005).

However, his method of the fewest line axial map generation, does not constitute an appropriate way to produce models for cities and regions. There are two primary factors, which prevent the application in a citywide and regional context. The first starts with the source of data and its definition of open space, a problem that the very initial convex space methodology already inherited. What to include and what to leave out in a graphical representation of the real world is left to the individual cartographer or researcher and forms core challenges in comparative cartography and map-making in general. This challenge is of particular importance when investigating suburban or rural areas. Suburban and rural areas often lack a continuous urban form and hence a given limitation for movement and visibility. Consequently, the definition of what can be considered an 'accessible open space' becomes vague. A problem that researchers are also facing in the context of developing countries exists as roads are often not solidified and boundaries between public and private spaces are less established. In these cases, an alternative could be to rely on other sources of geographic data of open spaces that follow precise definitions. Such sources are for example governmental agencies for cartography, geodesy and planning or volunteered geographic information, both of which have precise definitions of what and how open spaces are mapped.

Computational time constitutes the second difficulty. With a rising number of mapped open space polygons and their vertices, the necessary computational time to generate the fewest line axial map increases as well. Turner et al. give an account of the computational time needed for their algorithm to compute fewest line axial maps. A model of the small town of Gassin took 119 seconds to compute and featured 5217 lines in its initially generated all-line map and 38 axial lines in the final result (ibid.). Thus, the computational process for an entire city or even a region, with far more than one million street segments will take significantly longer[2]. While theoretically the algorithm could run for any time needed, in praxis this is limited by the software design dealing with large data sets. Currently the most commonly used software for this is depthmapX. Initial tests using the software on large urban systems generating fewest line axial maps have consistently produced application crashes. Varoudis et al. state the maximum number of segments that can be computed by depthmapX as <1.500.000 (2013), resulting in an axial line map of approximately 15000 lines. This makes an automated generation of axial lines for a metropolitan or regional system at the time not possible.

2 The total number of axial lines in cities with a population of 300,000 can range between 10,000 and 15,000.

2.1 ROAD-CENTRE LINES AS ALTERNATIVE FOR SEGMENT MAPS

Initially, the focus of axial line maps was to have a tool that allowed understanding complex urban systems in a simplified comparable manner. Over time the primary use of this morphological descriptive tool was to be found in investigations into the deep relation between human behaviour and space. From the development of the methodology, throughout the last 30 years, researcher have consistently found correspondence between the topological relationships of spatial systems and pedestrian movement (Hillier et al., 1993; Penn et al., 1998; Desyllas and Elspeth, 2001; Hillier and Iida, 2005) as well as vehicular movement activities (ibid.; Turner, 2005; Law and Versluis, 2015; Serra, Hillier and Karimi, 2015) and even global transportation networks (Hanna, Serras and Varoudis, 2013). This is particularly the case since the introduction of ASA in space syntax as an extension of axial analysis (Turner, 2001). The emphasis thus shifted from a theory and tool to analyse spatial configurations to one of predicting the potential of human behaviour in the form of movement and flows. Four studies focus on alternatives that constitute possible models for an analysis of movement and flows in the build environment: The pioneering work by Thomson (2003), Dalton et al. (2003), Turner (2005, 2007) and following up on these studies most recently the work by Dhanani et al. (2012). All authors investigate the possible application of different types of so called road-centre line data. The reasoning is that their approach relies on replacing a segment map, which is used in angular segment analysis rather than the in traditional axial line model the SM is based on. This study will follow the path taken by the above named researchers and base the comparison on a segmented axial line model, rather than emulating an axial line model, which inevitably will later be segmented in order to perform ASA.

Road-centre lines ideally represent the geographic centre of the public rights of way network, a transportation network of all paths on which the public have a legally protected right to pass and re-pass. These transportation networks are based on vector line information and can be generated through a variety of GIS methods such as automated processes of on ground collected GPS data, generative processes based on cadaster boundary data or manual tracing of roads on aerial photographs. In a subsequent step, additional information can then be attributed to this line information such as road names, road type, travel direction, road geometry information as well as a large variety of other possible attributes.

This makes road-centre line maps a powerful tool for a variety of GIS based applications. The ones applied the most are transportation modelling and navigation routing. Road-centre line data was first provided by local governments, such as the TIGER[3] data set by the United States Census Bureau or the ITN[4] by the British Ordnance Survey, as well as commercial companies, such as the Dutch Company TeleAtlas[5] or American-based Company Navteq.[6] The latter provides mainly line-based data for navigational systems. With the rise of the Internet and Web2.0[7], publicly accessible road centre-line information became largely available through different sources. The most predominant sources are Google maps and Bing maps, both available under restricted license for non-commercial usage. In contrast to governmental and proprietary based information with restricted license stands volunteered geographic information (VGI). VGI describes all geographic data, which is created, assembled and disseminated voluntarily by individuals (Goodchild, 2007). Open source VGI projects such as OpenStreetMap (OSM) and MapQuest are available under a GUP license and hence freely accessible to anybody. Due to the increasing number of online participants all over the world these projects are on the rise and establish a commercially as well as academically meaningful alternative.

3 TIGER is an acronym for Topographically Integrated Geographic Encoding and Referencing and an American based format used by the United States Census Bureau to describe land attributes such as roads, buildings, rivers, and lakes, as well as areas such as census tracts. The TIGER format forms a base for the US part of the OpenStreetMap project.

4 The Integrated Transport Network, is part of the OS MasterMap and a format provided by the United Kingdom governmental Ordnance Survey.

5 TeleAtlas is since 2008 wholly owned by navigation system company TomTom.

6 Navteq is since 2011 fully merged into NOKIA.

7 Web 2.0, is a term describing the state of the Internet as a collaboration focused information platform, where the user produces content. The term is set against Web 1.0, where content was provided as ‚ready-to-use' and no interaction with the user was aimed (O'Reilly, 2005).

In the context of space syntax analysis 2003, Thomson (2003) pioneered when proposing to make use of street networks. His study focuses on theoretical and technical problems based on the model construction rather than an investigation on how different models effects the analysis. In the study, he highlights possibilities of generalizing road networks. Simultaneously Dalton et al. propose to make use of TIGER data and present initial results of their analytical work (2003). TIGER is a data format only used in the United States providing road-centre line information among other geo-referenced spatial data. Dalton conducts a fractal analysis and compares a TIGER dataset with a traditional hand-drawn axial map of Downtown Atlanta, US. He highlights differences in the results of both models and concludes that the result is caused by the very different representation of space. While a long linear avenue with adjacent side streets is represented by one long axial line in a traditional axial line map in the TIGER dataset road centre-lines are segmented by nature and have a node at each intersection (this is the case for any road centre-line map). Any topological investigation would thus lead to a highly skewed outcome. Moreover, Dalton raises the theoretical problem of radii, emphasising the need for a 'relativisation' due to the differences within each system (ibid., p.9). While Dalton did not propose a solution to the problem his argumentation led to a series of investigations by Alasdair Turner.

In his study from 2005 Turner presents a methodology that overcomes this problem of segmentation and 'relativisation' by drawing on advantages of space syntax applying ASA to road centre-line maps in combination with a segment length weighted algorithm. The results of his 2005 and 2007 study indicate that metric radii in combination with weighted choice measures present not only a suitable alternative to SM models but, in fact, generate better correlations with flow data in the tested case studies. Turner emphasises that his measure holds configurational information while incorporating plausible cognitive and physical constraints (2007, p. 553). Turner's findings are reasonable since road centre-line maps are fundamental representations of the accessible – rights of way – movement network and incorporate more detailed angular information than axial line models.

Dhanani et al. (2012) follow Turner's findings and conduct a comparative study of an axial line model and two different types of road centre-line based models. As mentioned previously, there are different sources for road centre-line maps. Dhanani et al. studies' focus on two very particular networks: the governmental ITN data set and the OSM VGI data. Their studies aim to understand whether a VGI-based data set constitutes a reliable alternative compared to governmental data sets in the light of space syntax analysis. Beside of Dalton's (2003) and Turner's (2005, 2007, 2009) work, there are no other comprehensive studies where space syntax measures are applied to governmental road centre-line data sets correlating results with empirical data. This is surprising as both of the studies rely either on the American TIGER data or the British Ordnance Survey data sets. The difficulty here is that governmental road centre-line maps are presented as a reliable and coherent source of data, yet, this is only true for information within one data set [8] and very little is being said about their comparability in an international context.

Differences occur between governmental data sets not only on an international level but also within countries. The British Ordnance Survey for example provides three different road centre-line data products: the OS MasterMap layer Integrated Transport Network (ITN) layer, the OS Open Roads layer and the Merdian 2 layer. All these data sets provide comprehensive road network information and are designed for routing and road network analysis, yet, their level of precision and coverage differs.[9] This means that the total amount of nodes and coverage of real world details such as roundabouts are not the same throughout the three data sets. More importantly such data sets are not available in every country. Germany, Italy and France–to name only some–do not provide freely accessible data sets. This is why, the question of comparability needs to be answered and investigated for each country individually and alternative sources

8 It shall be noted that errors do occur in governmental data sets as well, but they usually follow a random distribution.

9 See http://digimap.edina.ac.uk/webhelp/os/osdigimaphelp.htm#data_information/os_products/os_open_map_local.htm for further information on the data sets and examples of their application.

need to be found. The lack of comparable data makes it difficult for international comparative approaches making use of such data sets, particularly in the context of space syntax.

2.2 ADVANTAGES AND DISADVANTAGES OF OSM DATA

In the light of this lack of comparable data, OSM data becomes more interesting as an appropriate alternative to a segment map representation, which, in theory, provides a comparable representation of space all over the world. OSM data is produced according to a guideline indicating the level of precision and the handling of particular situations such as divided highways, roundabouts, intersections or bridges (OpenStreetMap Wiki contributors, 2016). This makes the data, in theory, globally comparable. However, differences in terms of data quality arise due to the nature of its production and its contributors' heterogeneous understanding of street networks.

Understanding such differences in quality is a non-trivial task in the realm of OSM data. There is a set of ISO standardized quality measures to assess the quality of map-based VGI (OSM) data. These measures are of particular interest for routing and navigation application, namely *positional accuracy and topological consistency* (Senaratne et al., 2016, p. 6) and thus for a space syntax application. *Positional accuracy* is a quantifiable value reflecting the difference between a mapped location and its real world location while topological consistency measures how well topological relations ('disjoin', 'meet', 'overlap' or 'equal') are mapped. A simple example for low positional accuracy would be a mapped intersection, of which the GIS location is 20 meter further in the North than in reality. An example for bad topological consistency of an intersection would be the case, in which two streets, which in reality are connected and should share a common node, would not do so in GIS. To evaluate the two mentioned quality measures it is necessary to compare the data set under investigation with the real world. This is usually done by comparing the VGI data with ground-truth data. Ground-truth means data that represents the respective exact location in reality. This is a theoretical value, rather than an actually achievable goal for most GIS data sets. GPS systems feature on average a positional accuracy of 6-10 metres to ground-truth. The ordnance survey MasterMap ITN data states its positional accuracy with 1 metre in urban and 6 metres in rural areas against ground-truth.

Throughout the past decade, several authors have conducted comparisons of volunteered geographic information with governmental as well as commercially produced geographic information (Flanagin and Metzger, 2008; Neis et al., 2010; Zielstra and Zipf, 2010; Ludwig, Voss and Krause-Traudes, 2011)[10] to measure their quality. In the context of road centre-line information the work by Mordechai Haklay was one of the first to evaluate the quality of OSM data (2010). Haklay used the British OS Merdian 2 road network as control measure to test OSM data quality, his findings indicated highest mapping qualities in urban and affluent areas and the lowest coverage in rural and poorer areas while positional accuracy ranges from over 70% to occasionally drop down to 20% (ibid., p.700). Overall OSM data covered 29% of England based on a network from March 2008. In a subsequent study conducted in October 2009 this percentage was already corrected to 65% of coverage (Haklay, 2009). This indicates a growth of the network coverage by 36% within one year. Another study by Neis et al. (2011) dealing with the case of Germany, compared the OSM network against the proprietary data set of TomTom (formerly TeleAtlas) and estimated a complete coverage of the German OSM data by the year of 2012. Moreover, already in 2011 the OSM data exceeded the topological consistency and completeness of the TomTom network by 27% including pedestrian path ways (ibid.). The continuous growth and its pace of the OSM data set, does not only make a coverage and quality assessment difficult, but indicates that it is only a matter of time that full topological consistency will be reached. The number of total users in the OSM community as well as their nodal contribution to the network shows a growth of the total user number to 2,9 million since the start of the project 2004 and gives insights in the pace of this process.

10 See Sehra et al. (2013) and Senaratne et al. (2016) for a comprehensive review of studies dealing with quality assessment of VGI data.

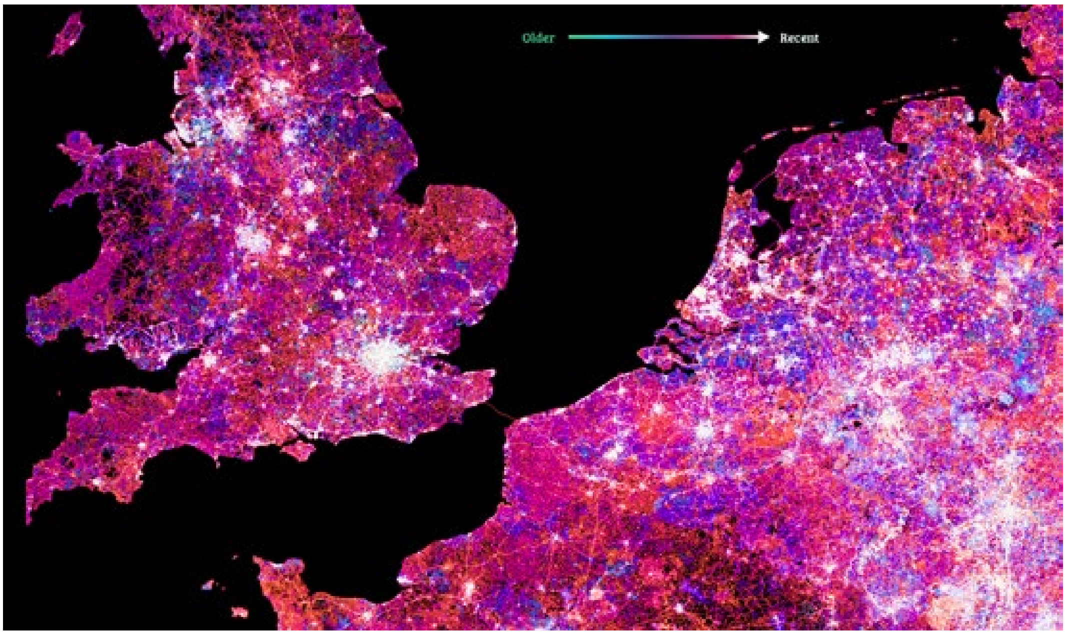


Figure 1 - Visualizing road updates. All roads shaded by how recently they have been updated by users. Older imports are in green and blue, while cities with strong and active communities and the effect of recent automated editing makes areas glow red. (2013) Source: https://www.mapbox.com/osm-data-report/ (retrieved on 1 August 2016)

Hakley et al. (2010, p. 11) investigate how many volunteers are needed to map an area thoroughly concluding that areas mapped by more than 15 contributors per square kilometre feature a very good positional accuracy of below 6 metres for resulting VGI data. In regard of the growing numbers of contributors this leaves us to expect an equal rise in topographic consistency and positional accuracy. An additional positive effect to the coverage of areas, beside the growing number of contributors, is the fact that governmental agencies increasingly provide their data for public usage. Likewise, are the American TIGER network as well as the AND Dutch road network fully implemented in the OSM network, aiding not only to the coverage but positional accuracy of the OSM data set. A visualised snapshot of the data and its topicality reveals updating intervals, as well as showing that Great Britain and Germany are part of the best-mapped countries of the OSM project (Figure 1). All of the above studies use ground-truth data for the evaluation of VGI quality. Still, such data is not available in every country and more difficulties for the assessment of VGI data arise due to the lack of ground-truth data for comparison (Senaratne et al., 2016, p. 6). To overcome this lack of ground-truth data, Keßler and de Groot (2013) propose a method to indicate quality of VGI via trust assessment models. Their approach is based on a trust assessment model of the independent contributions in an OSM data set. Albeit presenting promising results, the methodology is at an early stage of development and does not propose an applicable method for the field. At the present stage, this leaves the research with as-good-as complete network for some countries with reasonably accurate precision, but a manual control of the entire data set by the researcher stays a necessity. With regard to future research the OSM will very likely constitute the most coherent freely available data set.

Dhanani et al. (2012, p. 30), assess the usage of OSM in space syntax to be problematic and describe the data as lacking 'of consistency [,...] accuracy and coverage'. Their study calls on researcher to rely on governmental data such as the British OS MasterMap ITN, yet, as mentioned earlier, as data is not accessible in every country and level of detail differs throughout different data sets, this approach remains unsatisfactory: The OS MasterMap ITN network covers only the vehicular network disregarding any path or street that is only accessible to pedestrians. The resulting vehicular centred spatial representation can therefore only be used to evaluate vehicular structures. Space syntax segment map representation on the other hand sees space through the eye of an individual moving in space and constitutes a sharp contrast to a vehicular only street network. There are also other difficulties within the ITN data set that render an ad hoc use impossible. Dhanani et al. note that the ITN network comprises all traffic management features including traffic islands, artificial cul-de-sacs or roundabouts (ibid., p.6). According to the authors, using such data creates a 'disjoint and fragmented network' particularly if a researcher is interested in other modes than a purely vehicular estimation. The usage of such data is not recommendable without any prior processing. Prior processing is also necessary for OSM data making it indispensable to develop a strategy to overcome said inconsistency and arrive at a comparable network for any given case.

3. OSM DATA STRUCTURES AND GIS SIMPLIFICATION PROCESSES

The following section gives an overview of the necessary components to create a road network based on OpenStreetMap data and the necessary steps of post processes to allow an application in space syntax ASA.

At present, OSM data sets are divided into four different elements: nodes, lines, surfaces and relations. For an ASA only line information is necessary, but not all of the available line information and categories are useful. The OSM wiki provides extensive accounts on all different key categories and their morphology (OpenStreetMap Wiki contributors, 2017), it is important for each researcher working with OSM data to make him/herself familiar with all categories and morphologies. Decisions about which category to exclude might differ for example in cities in developing countries. The following steps should to be considered as a general guidance: For the purpose of network analysis only components with the key highway=* shall be used. This key defines any kind of road, street or path and their respective importance in the network hierarchy (from the most important 'motorway' to the least 'service') and, thus, gives a good account of the rights of way network. The following list assess which are recommendable to be included in a network for an application in ASA: highway=motorway; trunk; primary; secondary; tertiary; unclassified; residential; *motorway_link; trunk_link; primary_link; secondary_link; tertiary_link; living_street;* pedestrian (ibid.). Particular care needs to be taken with the key pedestrian as it includes pseudo polyline information of squares and these need to be cleaned and subsequently broken into individual segments. Other sub keys such as highway=*service; path or bridleways* can be included but are not recommended, as they are of very small scale and might otherwise be eradicated in a subsequent simplification process.

With a view to this selected data there are three main difficulties that occur when applied in a space syntax context.

1. Topological inconsistency occurs if street segments are supposed to share a connecting node but due to positional inaccuracy fail to do so. This is often the case at intersections of different contributors. Even a small gap between two nodal ends of 1 cm can create a network fragmentation. It is, therefore, necessary to process and clean the data from these inconsistencies.

2. Traffic management components are network details that are necessary for vehicular traffic management but have no immediate impact on cognitive route decision-making. Such details are for example roundabouts, small traffic islands or motorway trunks. Ideally roundabouts are simplified into simple intersections whereas meandering trunk links are represented by single links. Moreover, this is also the case with regard to dual line representations. Space syntax analysis is a non-directional approach in the sense that the possible travel directions are not taken into consideration and each space is treated as equally accessible. A dual line representation constitutes only a reasonable option if directions are taken into consideration. Hence, the model needs to be cleaned from said dual line representations.

3. Redundant or excessive nodal information are often problematic when using OSM data. Although the OSM guide notes that nodes should be used in an economic manner, contributors often have different interpretations of what 'economic' means. This is particularly the case for curved roads, but also occurs on straight lines. Ideally each street is simplified to its fundamental segment.

In order to overcome these difficulties a series of GIS algorithms have been developed. The following proposed solutions are employing the GIS software ArcGIS Desktop 10.2 from Esri. I employ ArcGIS because it is the only software that provides solutions for all three said difficulties. At present, only a few of the solutions presented here can be achieved with open source GIS software packages. Due to the scope of this paper only a brief description of the applied core functionalities will be given. Figure 2 shows a workflow diagram for the proposed solutions, while Figure 3 gives an illustration of each obstacle and its favoured solution after the application of the simplification method presented here.

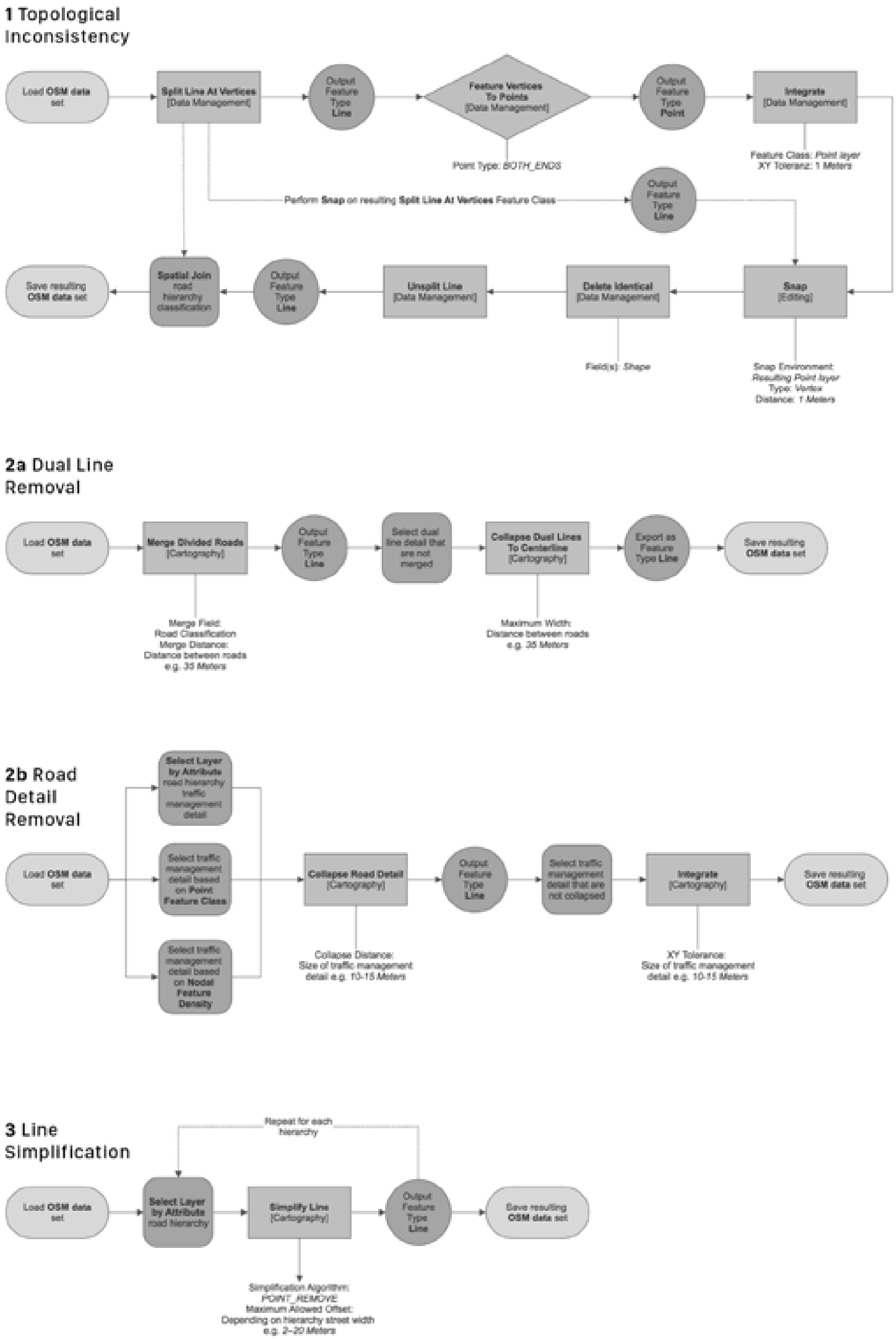


Figure 2 - Workflow of ArcGIS tools and algorithms to solve: 1. topological inconsistency; 2a. dual line removal; 2b. road detail removal and 3. line simplification.

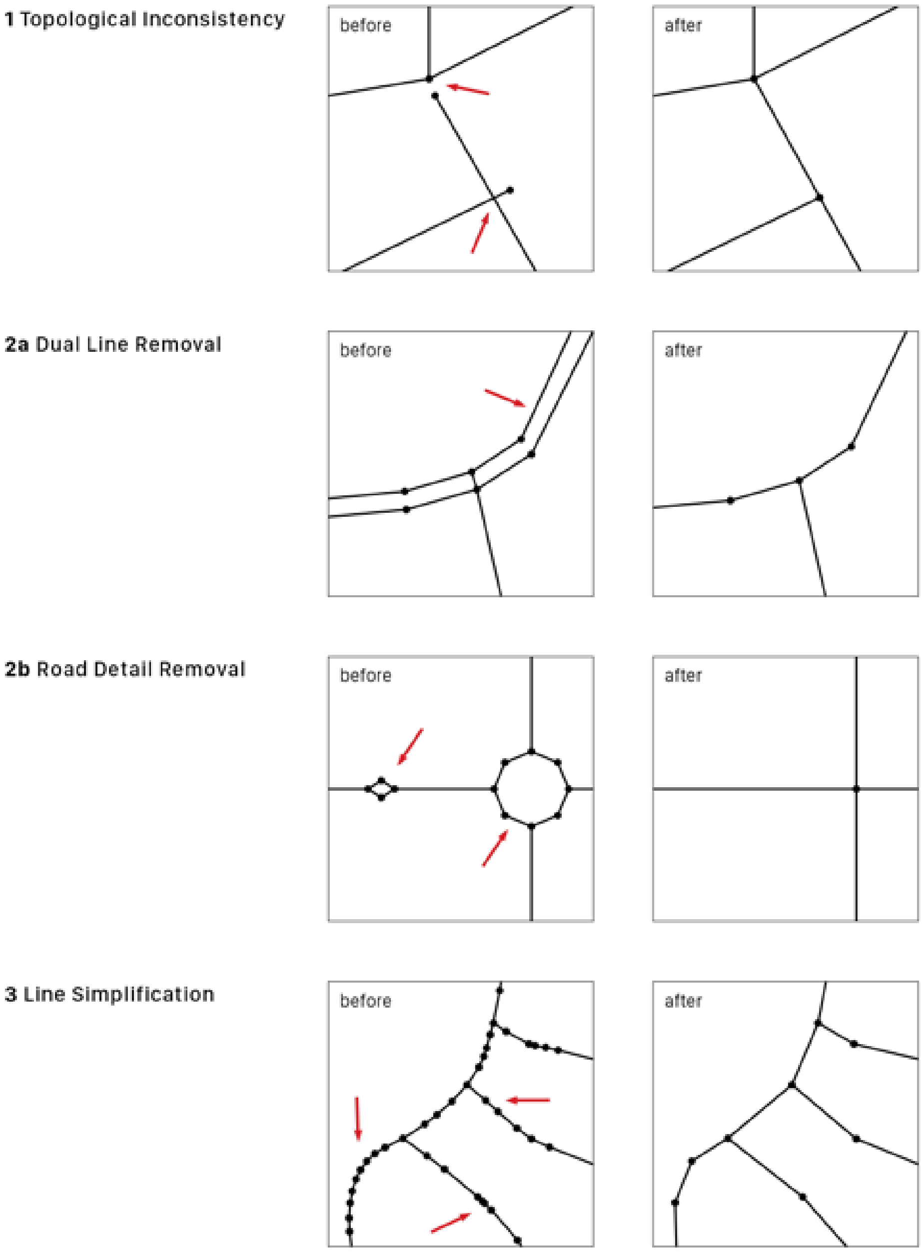


Figure 3 - Illustration of each difficulty found in OSM data: 1. topological inconsistency; 2a. dual line removal; 2b. road detail removal and 3. line simplification as well as the condition after application of the simplification method.

1. Starting with the approach of solving topological inconsistency (Figure 2:1), it should be mentioned that a lack of network information, such as entire missing streets cannot be solved through automated processing and that the OSM data needs to be carefully checked by the research prior to any post-production. More so, this is a strategy to overcome small inconsistencies that are difficult to identify manually. The proposed process reconnects topological inconsistencies by a given tolerance distance and in a subsequent step merge segments that can be considered as independent streets (from intersection to intersection) together. This will leave the researcher with a street network of real segments and consistent topological information. The two core ArcGIS functionalities the workflow is based on are *'integrate'* and *'unsplit'*.

The *integrate* tool is applied to extracted nodal information, rather than the actual line information, to overcome misalignment at intersections. Integrate maintains the integrity of shared nodal feature information by making features coincident if they fall within the specified x, y tolerance. Features that are considered identical or coincident are merged. In a subsequent step the newly generated nodal point information is used as a basis for a snap command of the initial street network. This will consequently connect lines, which feature topological inconsistencies, at a new point based on the location of their nodal line ends.

The unsplit tool is then applied to the now topological consistent line network. The aim is here to aggregate single part line features into multipart features in order to arrive with continuous street segments. Unsplit merges lines that have coincident endpoints. This can be done by relying on any given attribute information or, as in this case, solely by geometric relationships. Merged lines are of particular importance with regards to further simplification processes.

2. The next difficulty is the existence of traffic management details and dual line representations in the data sets (Figure 2:2a & 2b). Not only do such details (roundabouts, traffic islands, etc.) create differences in angular movement, while the general journey direction stays the same, but more importantly they increase the total number of journeys (dual line highways) and skew analytical results towards an emphasis of such details. Especially in the light of none directed centrality analysis dual lines make little sense. This could be negligible if traffic management details were normally distributed throughout the street network. However, this is, not the case with most examples and particularly not with inter-city and regional scales. There are four main ArcGIS components, *'merge divided roads'*, *'collapse dual lines'*, *'collapse road details' and 'integrate'* that help to remove such dual lines and reduce low-level street network complexity.

   The *merge divided roads* is an algorithm that merges road segments, which are parallel along a significant distance into a single centre line. The merging process is based on common attributes that can be computed on the basis of the initial highway keys. It is fundamental that the merge field parameters are established properly to avoid conflicts during the process. The divided roads algorithm can be applied to entire data sets and maintains topological relations with adjacent streets.

   The *collapse dual lines to road centerline* is an algorithm designed to derive with centre lines from a base of street perimeters. It is, therefore, a less sophisticated form of simplification and it is not recommended to perform the algorithm on large datasets including multiple-lane highways with interchanges, ramps, overpasses and underpasses. In individual cases where the merge divided roads tool does not arrive with satisfactory results, the *collapse dual line to road centerline* tool can form a useful alternative.

   The *collapse road detail*, on the other hand is an algorithm that depicts small road segment details and open configurations that interrupt the general trend of a road network and collapses or replaces them with a simplified feature. The collapse distance on which the tool performs is defined by the maximum size of the largest road detail and can differ for each model. If the *collapse road* detail tool does not solve or remove some of the details the integrate tool explained earlier constitutes an appropriate alternative. Particular care needs to be taken when using integrate on road details as it can impact the topological consistency of the data and should hence not be performed on entire data sets but single cases.

3. Line simplification is usually applied when segment records feature far more data than necessary for computer analysis or visual representations (Figure 2:3). In the case of space syntax and the use of VGI street networks this poses a conceptual question aside of excessive data. While road-centre lines depict the centre of the road an axial line (as base for a segment map line) is based on the longest line of sight. A generic street usually features a much larger field of vision than that of a single line. While axial lines fundamentally connect convex spaces these lines naturally pervade more than one space

at once. Road-centre lines on the other hand simply represent the centre of the road and, therefore, feature excessive angular information that does not impact the field of vision or accessibility and, thus, has no effect on the actual movement in space. This is why, a removal of such road details should be based on the field of vision of each street, i.e. the street width. Since road-centre lines give a precise account of the centre of each street segment a simplification process should allow the newly generated feature to deviate to at least the extent of the field of vision. Such processes can be performed by the Douglas-Peuker Algorithm (DPA) (1973). The DPA is broadly considered to deliver the best perceptual representations of the original segment and generates new segments based on a deviation tolerance. In ArcGIS this can be done by applying the *simplify line* tool.

The *simplify line* tool reduces and removes redundant nodes of line features. Among others, when applied with the POINT_REMOVAL functionality it employs the DPA. The aim of the algorithm is to extract the essential segment form based on a previously selected off set tolerance. The strength of the algorithm is its reproducibility and process speed, and that it arrives at the same solution to the same given problem.

If the above steps of the methodology are followed the simplified version of a road-centre line map (SIMP) looks visually as well as topologically much closer to an axial line representation.

4. MODEL EVALUATION METHODOLOGY

In order to test if the theoretically laid out version of a simplified OSM network (SIMP) constitutes a comparable alternative to a segmented axial line map and is, thus, suitable for the purpose of analysis of different scales and very large ones in particular the model will be analysed and correlated with results from an ASA of a segment map, ITN and OSM model. The comparison extends and the builds on methodologies by Eisenberg (2007), Turner (2007) and Dhanani et al. (2012).

Eisenberg (2007, p. 5) focused on comparison of different axial line models for the same cities. The different models that Eisenberg compares are developed as a by-product of variations in analytical scales (pedestrian, bicycle and vehicular) and variations in the detail of the base information used for the production of the axial line maps. Eisenberg highlights that three indicators are of interest for a comparison. First, the impact of base map scales; Second, different levels of detail; And, third, different city morphologies (ibid., p.5). All aspects are directly transferable to the different network models previously introduced. Eisenberg's findings suggest that the analysis should focus on 'rank correlation measures' in order to have a meaningful comparison (ibid., p.8). Eisenberg's 'rank correlation measures', are applicable to every kind of network representation. This measure simply compares values and their respective rank within the data set. With Eisenberg's measure an appropriate method for the aimed analysis is established where numbers of lines differ significantly and the resulting values do not form a comparable unit.

In addition to 'rank correlation' this comparison will draw on the methodology of Turner (2007). Turner proposed an *angular* based analysis in combination with segment length-weighting and the introduction of a *metric length* based *radius.* While an angular based analysis incorporates the cognitive dimension of route choices, the reasoning behind a *segment length-weighting* is to overcome the large differences in segment numbers between the different representations (ibid., p.541). Turner shows how his propositions are an advancement for space syntax analysis in general and in the context of road-centre line networks in particular.

Finally the above proposed methods will be merged with a methodology by Dhanani et al. (2012). Dhanani et al. conducted a comparison of road-centre line networks against axial line models using a general description of the network characteristics followed by a topological and metric step depth analysis from the most central segment. Although the outcome of the topological step depth showed interesting results the application of topology on a road-centre line network remains inappropriate as road-centre lines topological information is highly skewed by its nodal information. The measure of topology in space syntax analysis is based on the cognitive and visual space in the sense that what is considered as one space in space syntax would result in several spaces in a road-centre line network. The analysis will only draw on the measure of metric step depth (MSD) for comparisons as MSD is not affected by nodal information.

In summary, the following comparison is based on four different road network models of the centre of the city of Leeds. The city of Leeds was selected because it features a variety of different network details such as motorways, traffic management details as well as local paths. The road network models are: the Ordnance Survey ITN network, the OSM network, a simplified version of the OSM (SIMP) and a segmented axial line model (SM). The ITN network and the OSM data are not simplified but instead used as they are provided by the organisations. Moreover, the ITN and OSM networks where controlled on topological consistence, yet, no irregularities were found. Some network categories, as those mentioned in the OSM data sections, have been removed from the OSM data set while traffic management details remained unchanged. The four models are compared in regard to their network characteristics and analysed on 14 different radii from 100 up to the entire system n[11] using angular segment analysis with segment length weighting. The models are analysed on closeness and betweenness centrality. The resulting structures of three exemplary scales are visually compared. Then, subsequent correlations are conducted using 'rank correlation measures'. To facilitate comparisons mean values of coincident segments of the ITN, OSM and SIMP with the SM model are plotted on each respective SM segment.

4.1 RESULTS

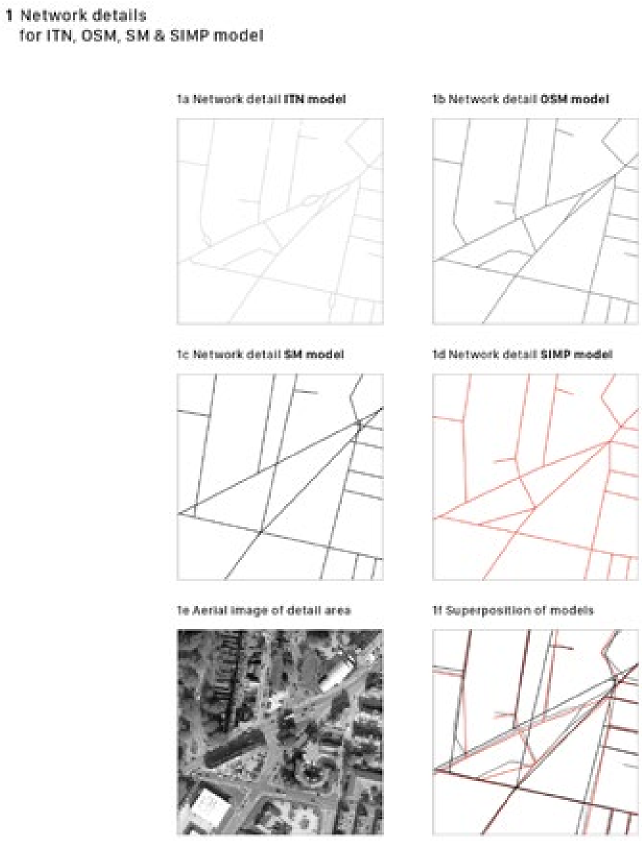


Figure 4 - Detailed section of different network models of ITN, OSM, SIMP and SM.

11 The applied scales are: *100, 150, 200, 300, 500, 800, 1300, 1800, 2500, 3200, 4100, 5000, 6100* and *n*.

Figure 4 shows a small section of each of the modelled areas. The section of the ITN network shows traffic islands as well as road interruptions. Some roads have significant angular turns just before their connection with the adjacent road. This is because for traffic management purposes rectangularity is preferred. In the light of angular segment analysis, Dhanani et al. (ibid., p.10) consider this preference an important aspect and the most detailed and 'optimal' account of the street network. The aerial photo of the area (Figure 4) shows that at this point a straight connection is a more reasonable account of the real world situation. Additionally, at the lower right there is a road divergence into two separate lanes. A noteworthy detail is also that roads, which could be considered as intersecting in reality do not share a common node in the road network, due to a 5-10 metre distance of their road-centre.

Statistics	ITN	OSM	Axial	SIMP
Segments	15049	9308	5072	3908
Total length (m)	283410	276388	240534	238848
Computation time (min)	14.31	4.49	1.21	0.44

Table 1 - Network characteristics for each model.

Table 1 highlights the network characteristics for the four models and how they differ numerically. The ITN network features the longest total network length with 283410 metres. This is particularly due to the several roundabouts and traffic management details within the model. The comparison of traffic management details with the length of the ITN and OSM networks enables a rough account of the effect on the length of the network. This account does not come to its fullest as the OSM network features streets and connections that are not represented in the ITN. The several multi line motorway roads, which are represented by a single segment in a segmented axial line and SIMP model cause a large difference of 40km of the ITN and OSM data in comparison to the segmented axial model. Comparing all networks, the difference in number of segments is striking. The ITN model has three times more segments than the segment map representation. This difference is due to the curved roads and roundabouts, which feature large numbers of segments in order to give precise accounts on the length of the lines. While this exemplifies the detailed account on angular changes in road centre-line networks, it also shows the inherent problem this data has when it comes to space syntax analysis. The computational time is O(n2) to the number of segments. Generally speaking, the ITN and OSM are similar in their measures and the difference in number of segments is as expected. With regard to the segmented axial line and SIMP model the question is whether the SIMP model, with 33% less segments, does also stores less information. The number differences can be explained by the 'cleaning' of intersecting spaces: Whenever three segments intersect with each other segmented axial line models tend to create clusters of very short segments. Additionally, when the axial line model is converted to a segmented axial line, stubs that fall over 40% of the line length are not removed and might also contribute to this difference. The SIMP model features almost the same length as the segmented axial line model pointing towards a similar degree of spatial representation.

1 Histogram of segment length distribution for ITN, OSM, SM & SIMP model

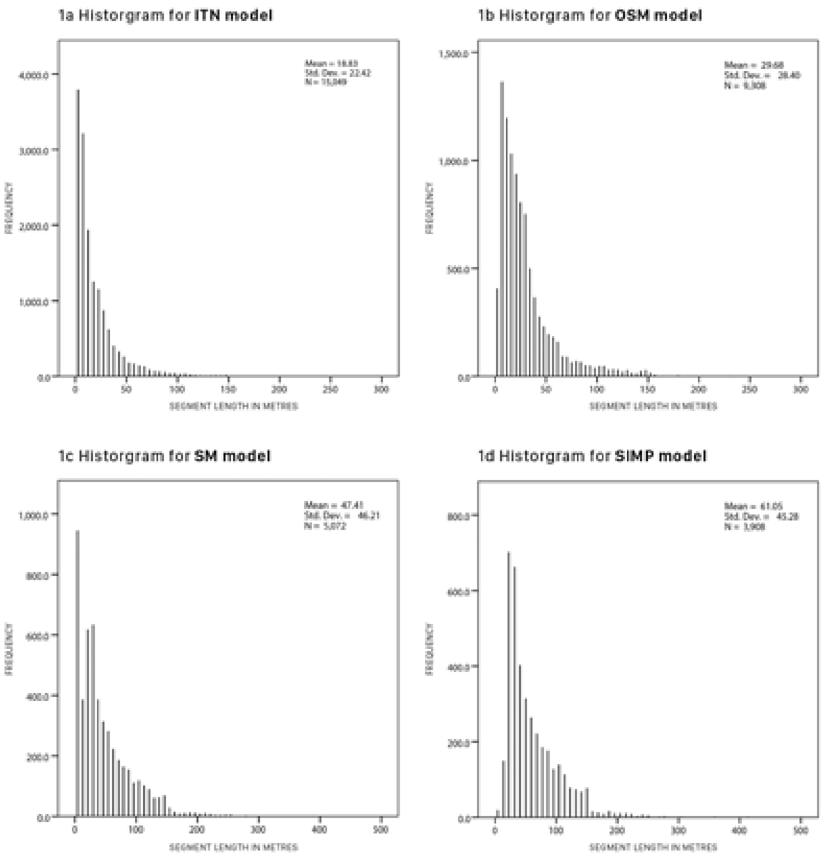


Figure 5 - Histogram of segment length distribution of each of the four models.

These observations become more apparent with a look at the histograms for segment length distribution for each m odel type. While the ITN network exhibits an even increase of segment length with declining frequency, the OSM shows an initial increase indicating fewer stubs and curve segmentation than the ITN. Moreover, the short line cluster effect of the SM model becomes visible with almost thousand segments in the range of approximately 1-10 metres. Contrarily, the SIMP model has a steep increase of frequency with a peak at a mid range of approximately 30 metres indicating less of a short line information. The simplification range used during the simplification process has an influence on this peak.

Dhanani et al.'s (ibid., p.25) study shows that differences between road centre-line network and axial line models are consistent in their appearance and concludes that the different models do not form a fundamentally different structure of the spatial configuration. In the next step I will compare the new SIMP model with this assumption. Figure 6 shows the number of segments for nine different radii where the maximum is 2,5km as this is the distance at which the entire system was captured (in other words n). For the four models, the total number of segments reached per metric distance increases in relation to the total number of segments. The semi-log plot highlights these similarities and differences, especially at lower scales. The SM and SIMP model, exhibit a similar development, while the OSM and ITN, which were initially similar, disperse towards growing metric distances and due to the increase of network details. Unlike the values for the central segment the curve for the edge segment shows a slightly uneven development. This becomes clearer in the semi-log plot of the data. Here, particularly the development around the scale of 500 metres unveils that there are underlying differences in the complexity of the models that might have an effect on the analysis.

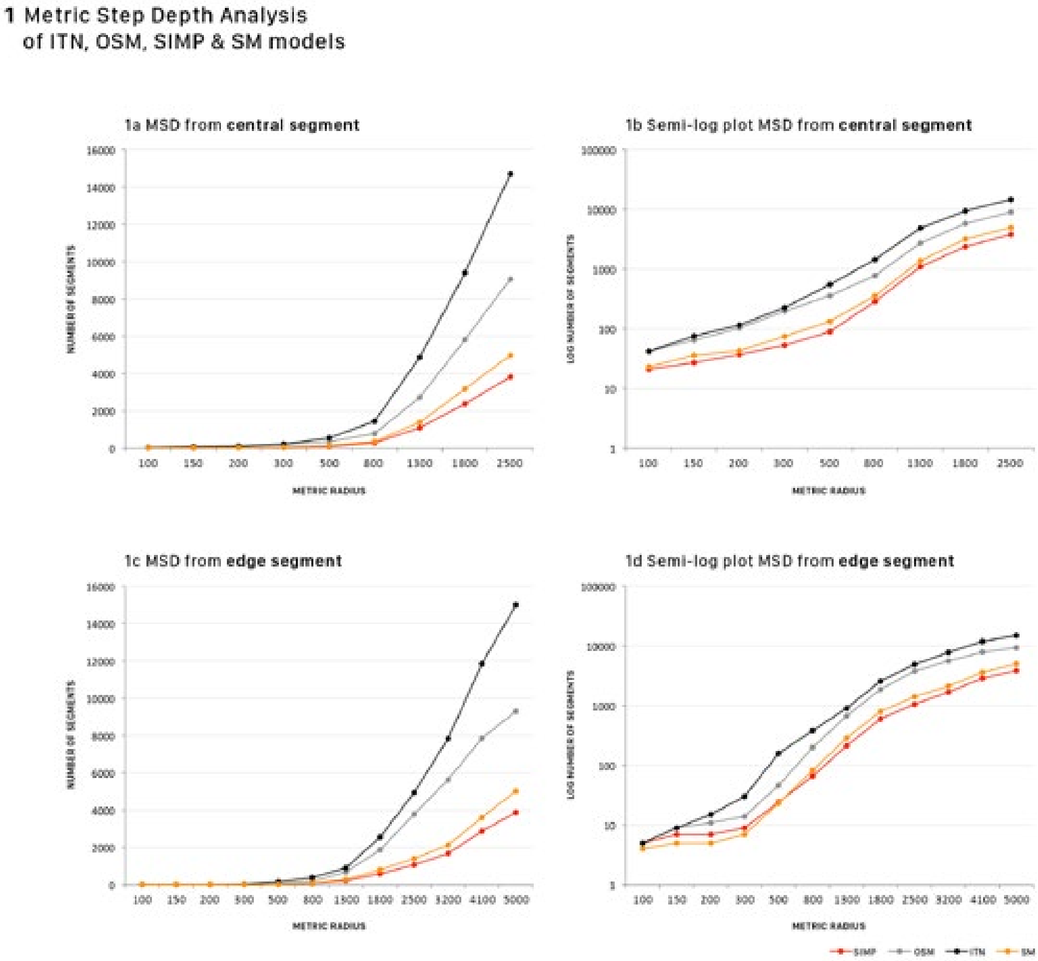


Figure 6 - 1a: Number of segments for different metric step depth from the most central segment for ITN, OSM, SM and SIMP models. 1b: Semi-log plot of the same data set. 1c: Number of segments for different metric step depth from an edge segment for ITN, OSM, SM and SIMP models. 1d: Semi-log plot of the same data set.

In order to arrive at a better and more detailed account of the impact of differences in the network morphologies, I conduct a comparison of betweenness and closeness centralities using a segment angular analysis with segment length weighting. The models are analysed on 14 different radii. The applied scales are; *100, 150, 200, 300, 500, 800, 1300, 1800, 2500, 3200, 4100, 5000, 6100* and *n*. Two of these scales, 800 and n, are visualised in order to understand the geographic distribution of differences. Figure 7 shows the results for betweenness centrality. Figure 8 shows the results for closeness centrality. The values of each figure are broken down using a quantile division. This is done to overcome significant outliers in the data sets that make a natural break highly skewed and the resulting maps illegible. These circumstances make it necessary to process the data in a GIS programme rather than applying the implemented symbologies of depthmapX.

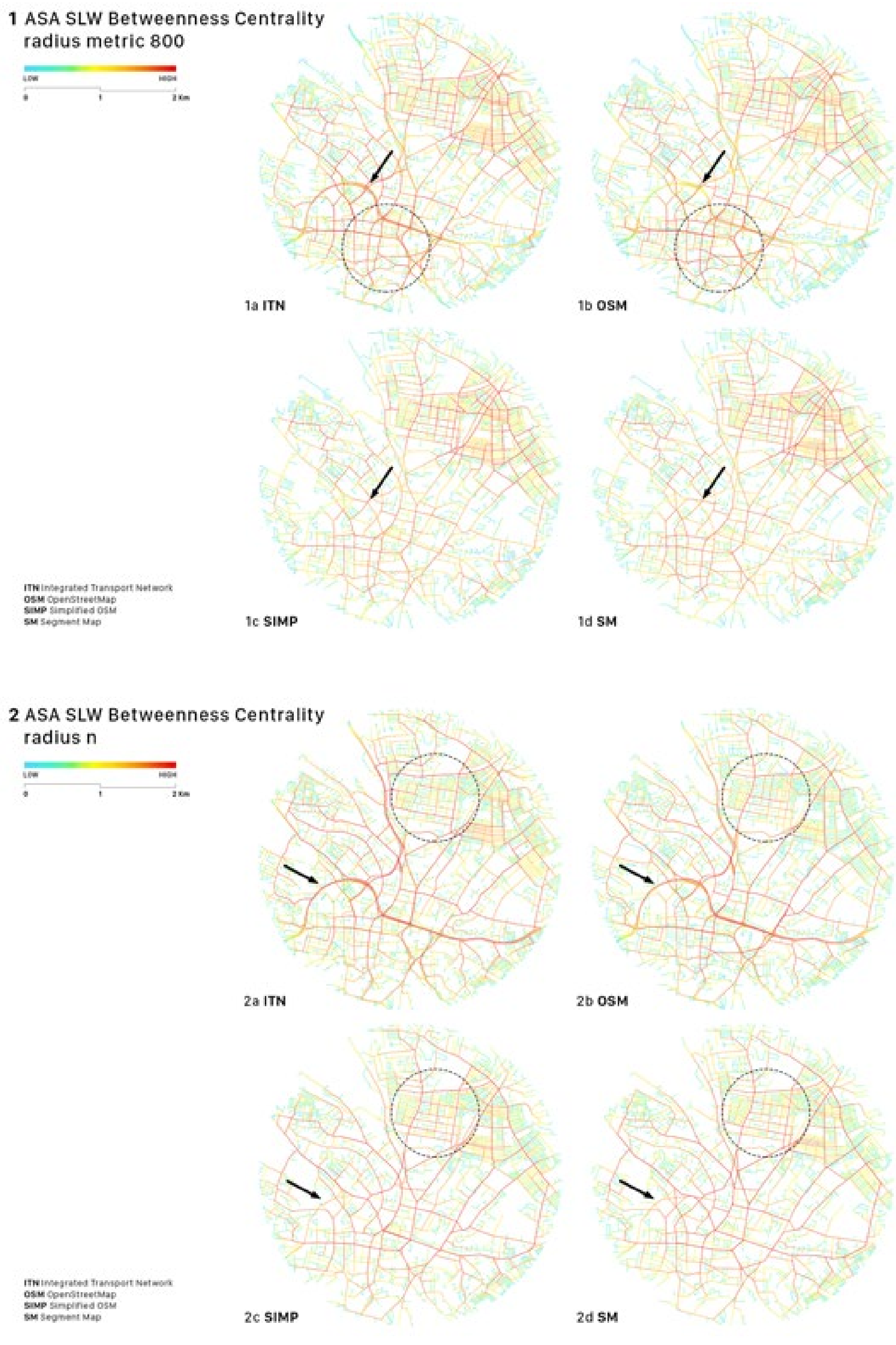


Figure 7 - ITN, OSM, SIMP, and SM models analysed on ASA SLW betweenness centrality on radius metric 800 (1) and radius n (2).

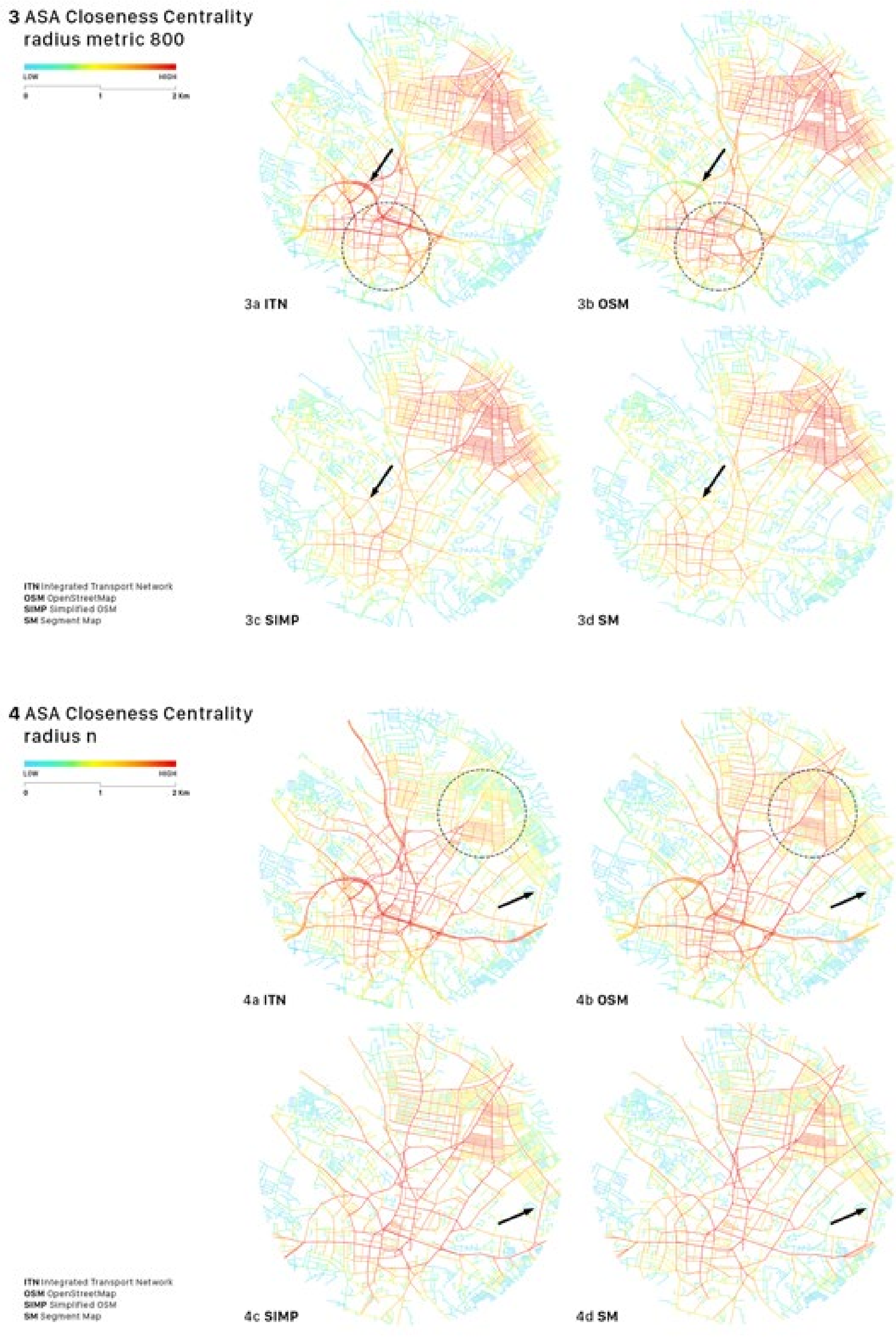


Figure 8 - ITN, OSM, SIMP, and SM models analysed on ASA closeness centrality on radius metric 800 (1) and radius n (2).

The results show that all models exhibit comparable patterns on all of the two visualised scales and both measures of betweenness and closeness centrality. This confirms the initial findings of Dhanani et al. (2012). However, similarities in the results were much stronger between the OSM network and the SM than they were between ITN and SM. Nominal segment differences appear to have a higher impact on betweenness centrality than on closeness centrality. Models with large numbers of short segments and high degree of precision, such as the ITN network, are, thus, more likely to be affected by outliers and unexpected clusters, than models with fewer short segments. Moreover, the ITN network shows high values on all scales in the motorway network. The SIMP model showed patterns that were visually stronger related to the SM model than to the ITN or OSM and more similar to the OSM compared with the ITN. This is rather unexpected as SM models are thought to be intrinsically different.

After getting an understanding of differences and similarities in the geographical distribution of the data between the different models and the SIMP model in particular a final analysis of the statistical extent of these observations is conducted. This will give an account of how models behave in comparison to each other across all scales. As elaborated before, the analysis draws on Eisenberg's proposed 'rank correlation measure'. To give a more detailed account, differently to Eisenberg, this analysis will compare all segments that are intersecting rather than only 10% of highest values proposed by Eisenberg (2007). This is done by plotting mean values of the ITN, OSM and SIMP on the SM model. The SM model is used as a base and comparisons are only conducted with streets whose middle point falls into a 10-metre distance of a SM segment. These middle points are then snapped to the closest segment and plotted on the SM model. If more than one street segment of an ITN, OSM or SIMP model falls into this category, their mean is calculated and plotted on the SM model instead.

Eisenberg's rank correlation is based on Spearman's Rank correlation (ibid.). Spearman's Rank correlation coefficient is generally used to identify and test the strength of a relationship between two sets of data. It tests if the relationship of both variables can be described by a monotonic function. Ideally, the SIMP model could predict the segmented axial line model by such monotonic function. In addition to this, a Pearson correlation will be conducted. Rather than correlating the different ranks of each variable, a Pearson correlation works with the actual values of the variables and measures their linear correlation. Both correlations provide a coefficient R2 indicating how related the variables are with each other. A coefficient of 1 indicates that the two models are identical. Any value below 1 describes the degree of difference. One can hence compare the differences between all models statistically and provide a correlation coefficient to describe the fitness of the SIMP model for the purpose of space syntax ASA. The analysis is based on 14 different scales for both space syntax measures of betweenness and closeness centrality. Figure 9 and Figure 10, show Pearson and Spearman correlations of ITN, OSM and SIMP compared with the segmented axial model and, subsequently, the same for all models correlated against the SIMP model.

Starting with Figure 9 the findings from the initial visual description becomes also statistically apparent. A first observation is that the Spearman rank correlation provides more consistent results across scales and measures with weaker differences and higher scores. The Pearson correlation on the other hand shows much stronger differences in the four data sets but features a significant outlier on the scale of 100 metres for closeness centrality. In regard to the single models the ITN model shows lower correlations across both Pearson and Spearman measures and on both betweenness and closeness centrality. Particularly interesting is the significant drop towards higher radii, with a lowest correlation of 0,56 on Pearson for betweenness and closeness. This increases at the Spearman's rank, however, the general tendency towards lower correlation at higher radii persist. In terms of the visual observations made earlier this is caused by traffic details and the strong representation of motorway features. The OSM and SIMP model on the other hand show very comparable correlation developments. An exception of this is the Persons correlation for betweenness centrality of the OSM model where similar to the ITN a sudden drop at higher radii is visible. The SIMP model correlates stronger across all measures with the highest scores of 0,983 for Spearman correlations of closeness centrality metric 1300 and 0,919 for betweenness centrality. Contrary to OSM and ITN the correlations for SIMP are very consistent.

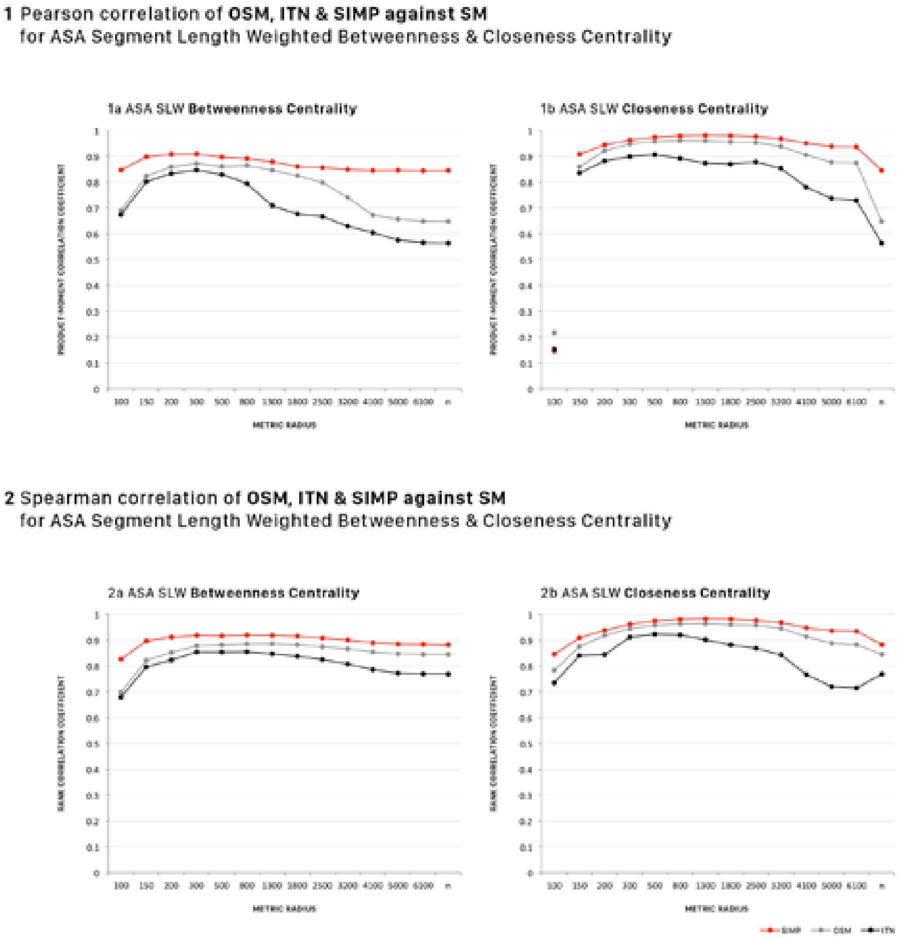
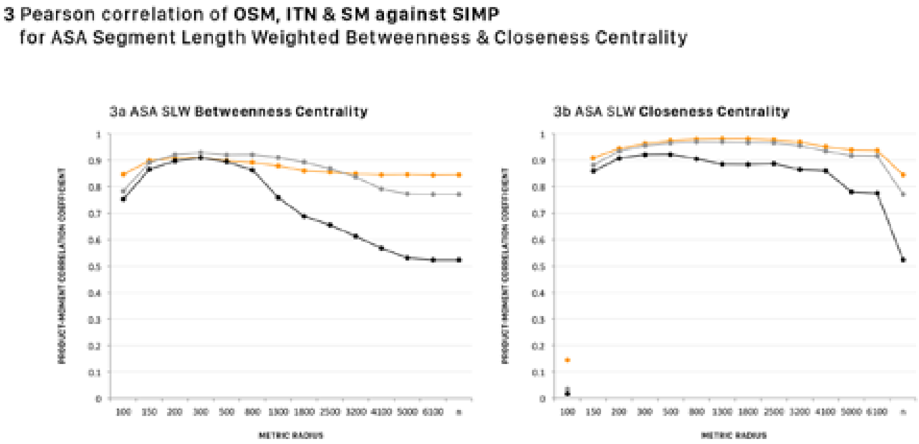


Figure 9 - 1: R2 of a Pearson correlation for ASA segment length weighted betweenness centralities (1a) and closeness centrality (1b) for 14 different metric radii (from 100 metres to n) for the three different network models SIMP, OSM and ITN against the SM model. 2: R2 of a Spearman correlation for ASA segment length weighted betweenness centralities (2a) and closeness centrality (2b) for 14 different metric radii (from 100 metres to n) for the three different network models SIMP, OSM and ITN against the SM model (left). Correlation is significant at the 0.01 level (2-tailed), N=3172.

Figure 10 shows the Pearson and Spearman correlations for 14 different scales and closeness and betweenness centralities. However, this time ITN, OSM and SM models are compared with SIMP. The general correlation developments are very similar to the ones we have observed previously, with a progressive drop of values towards higher radii. Interesting is at this point how ITN and OSM behave compared to the SIMP model. While the ITN networks shows a slightly weaker correlation, the OSM correlates much stronger. This was on one hand an expected result, as the SIMP model is entirely based on the OSM. On the other hand in the light of the overall comparison it seems as if the simplification process brought the simplified OSM model much closer to the segmented axial line representation than expected.



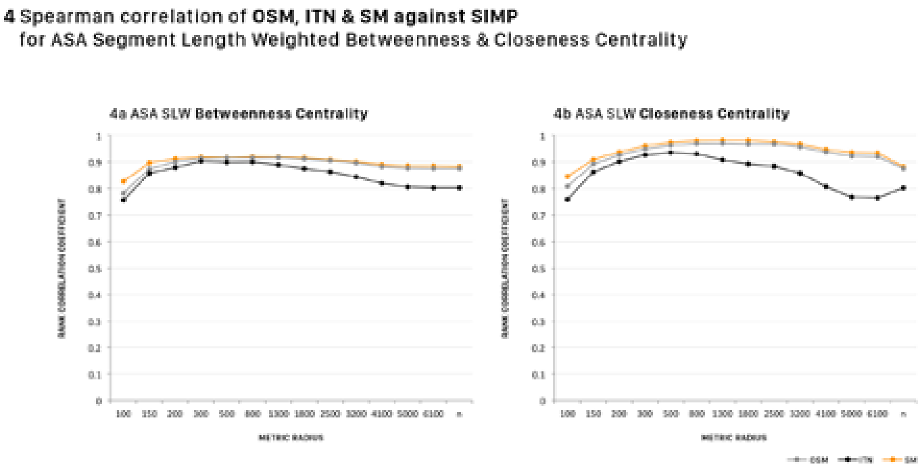


Figure 10 - 3: R2 of a Pearson correlation for segment length weighted betweenness centralities (3a) and closeness centrality (3b) for 14 different metric radii (from 100 metres to n) for the three different network models SM, OSM and ITN against the SIMP model (left). 4: R2 of a Spearman correlation for segment length weighted betweenness centralities (4a) and closeness centrality (4b) for 14 different metric radii (from 100 metres to n) for the three different network models SM, OSM and ITN against the SIMP model (left). Correlation is significant at the 0.01 level (2-tailed), N=3172.

These differences become more apparent with regard to a log-log scatterplot of betweenness and closeness centrality of the global scale *n* (Figure 11). The diagram shows a log-log scatterplot of each of the measures allowing a visual comparison of outlier distribution within each data set. The more dispersed the values are the less they correlate while linear consolidation implies stronger correlations. This is clearly visible for the log-log plot of axial and SIMP while both other models show stronger dispersion. The ITN model shows outliers across the values from low to high, which is particularly the case for closeness centrality. To summarize, the results

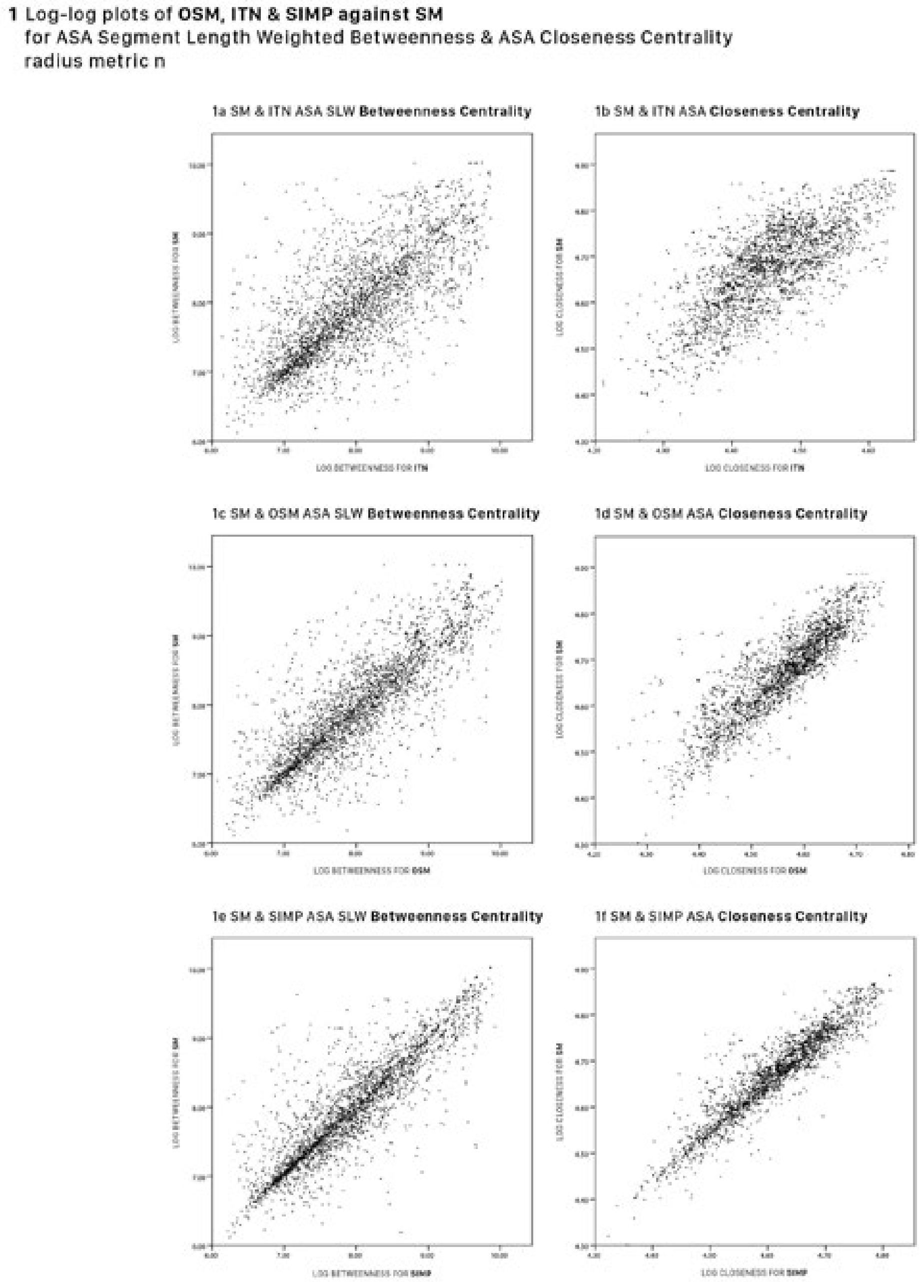


Figure 11 - Log-Log plots for the SM model compared to ITN, OSM and SIMP respectively for ASA SLW betweenness and ASA closeness centralities on radius n.

show that the four models differ especially in terms of the number of short length segments. This difference can be described by an exponential relation and has a significant impact on the computational time needed for the analysis. The results of the metric step depth analysis confirm the findings of Dhanani et al. (2012) and show that all models share a similar complexity in terms of their nodal distribution. However, the analytical space syntax analysis showed that, albeit, there is a similar distribution in the data in general the geographic location of these differences has an impact on the results. The ITN network is strongly influenced by its emphasis on vehicular movement and traffic management details. This makes it less comparable to the segmented axial line model than the OSM model or the SIMP.

5. CONCLUSIONS

Concluding, this paper elaborated the fitness of OSM data in space syntax analysis, it proposed an ArcGIS simplification workflow and presented the theoretical reasoning behind the method. The final fitness tests showed that the simplified OSM network (SIMP) exhibits very strong similarities with the traditional segmented axial line model across all investigated cases. It features the topological and angular information of the OSM network with the simplistic representation of a segmented axial line model. This is rather surprising, because the alterations in the model are mainly based on segment nodal reduction and minor topological alteration. The Pearson and Spearman correlation analysis showed that the SIMP model is in fact stronger related to the segmented axial model than to the OSM model. The strong similarity between SIMP and segmented axial also poses question to weather axial line models are such intrinsically different representations.

Overall the findings suggest that a simplified OSM network forms an appropriate model for space syntax analysis, particularly in the light of regional investigations where the production of an axial line model is not a feasible option.

REFERENCES

Batty, M. and Rana, S. (2004) 'The automatic definition and generation of axial lines and axial maps', *Environment and Planning B: Planning and Design,* 31(4), pp. 615–640. doi: 10.1068/b2985.

Carranza, P. M. and Koch, D. (2013) 'A Computational Method for Generating Convex Maps', in Kim, Y. O., Park, H. T., and Seo, K. W. (eds) *Proceedings of the Ninth International Space Syntax Symposium.* Seoul: Sejong University, p. 064: 1-11. Available at: http://www.diva-portal.org/smash/get/diva2:679788/FULLTEXT01.pdf.

Dalton, N. S., Peponis, J. and Dalton, R. (2003) 'To tame a TIGER one has to know its nature: Extending weighted angular integration to the description of GIS road-centerline data for large scale urban analysis', *4th International Space Syntax Symposium,* pp. 1–10.

Desyllas, J. and Elspeth, D. (2001) 'Axial Maps and Visibility Graph Analysis. A comparison of their methodology and use in models of urban pedestrian movement', in *Proceedings of the Third International Space Syntax Symposium.* Atlanta, USA, p. 27.1-13. Available at: www.ucl.ac.uk/bartlett/3sss/papers.../27_desyllas.pdf.

Dhanani, A., Vaughan, L., Ellul, C. and Griffiths, S. (2012) 'From the axial line to the walked line: Evaluating the utility of commercial and user-generated street network datasets in space syntax analysis.', *Proceedings of the Eighth International Space Syntax Symposium.* Santiago de Chile: PUC, pp. 1–32. Available at: http://discovery.ucl.ac.uk/1308812/.

Douglas, D. H. and Peucker, T. K. (1973) 'Algorithms for the Reduction of the Number of Points Required for Represent a Digitized Line or its Caricature', *Cartographica (The Canadian Cartographer)*, 10(2), pp. 112–122.

Eisenberg, B. (2007) 'Calibrating axial line maps', in *Proceedings of the Sixth International Space Syntax Symposium.* Istanbul, Turkey, pp. 1–14.

Flanagin, A. J. and Metzger, M. J. (2008) 'The credibility of volunteered geographic information', *GeoJournal,* 72(3–4), pp. 137–148. doi: 10.1007/s10708-008-9188-y.

Goodchild, M. F. (2007) 'Citizens as sensors: The world of volunteered geography', *GeoJournal,* 69(4), pp. 211–221. doi: 10.1007/s10708-007-9111-y.

Haklay, M. (2009) *OpenStreetMap and Ordnance Survey Meridian 2 – Progress maps.* Available at: https://povesham.wordpress.com/2009/11/14/openstreetmap-and-ordnance-survey-meridian-2-progress-maps/ (Accessed: 10 June 2016).

Haklay, M. (2010) 'How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets', *Environment and Planning B: Planning and Design,* 37(4), pp. 682–703. doi: 10.1068/b35097.

Haklay, M., Basiouka, S., Antoniou, V. and Ather, A. (2010) 'How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information', *The Cartographic Journal,* 47(4), pp. 315–322. doi: 10.1179/000870410X12911304958827.

Hanna, S., Serras, J. and Varoudis, T. (2013) 'Measuring the structure of global transportation networks', in Kim, Y. O., Park, H.-T., and Seo, K. W. (eds) *Ninth International Space Syntax Symposium.* Seoul: Sejong University.

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

Hillier, B. and Iida, S. (2005) 'Network and psychological effects: a theory of urban movement', (1987), pp. 475–490. Available at: papers2://publication/uuid/51712050-C088-4BAE-B7A0-6A4372B28C46.

Hillier, B., Penn, A., Hanson, J., Grajewski, T. and Xu, J. (1993) 'Natural Movement: or, configuration and attraction in urban pedestrian movement', *Environment and Planning B: Planning and Design,* 20, pp. 29–66.

Keßler, C. and de Groot, R. T. A. (2013) 'Trust as a Proxy Measure for the Quality of Volunteered Geographic Information in the Case of OpenStreetMap', in *Lecture Notes in Geoinformation and Cartography,* pp. 21–37. doi: 10.1007/978-3-319-00615-4_2.

Law, S. and Versluis, L. (2015) 'How do UK regional commuting flows relate to spatial configuration ?', in Karimi, K., Vaughan, L., Sailer, K., Palaiologou, G., and Bolton, T. (eds) *Proceedings of the 10th International Space Syntax Symposium.* London: Space Syntax Laboratory, The Bartlett School of Architecture, University College London, p. 74:1-74:21.

Ludwig, I., Voss, A. and Krause-Traudes, M. (2011) 'A Comparison of the Street Networks of Navteq and OSM in Germany', in Geertman, S., Reinhardt, W., and Toppen, F. (eds) Berlin, Heidelberg: Springer Berlin Heidelberg (Lecture Notes in Geoinformation and Cartography), pp. 65–84. doi: 10.1007/978-3-642-19789-5_4.

Neis, P., Zielstra, D., Zip, A. and Strunck, A. (2010) 'Empirische Untersuchungen zur Datenqualität von OpenStreetMap - Erfahrungen aus zwei Jahren Betrieb mehrerer OSM-Online-Dienste', in *Angewandte Geoinformatik 2010.* Salzburg, Austria, p. 6. Available at: http://www.vde-verlag.de/proceedings-en/537495055.html.

Neis, P., Zielstra, D. and Zipf, A. (2011) 'The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011', *Future Internet,* 4(1), pp. 1–21. doi: 10.3390/fi4010001.

O'Reilly, T. (2005) What Is Web 2.0: *Design Patterns and Business Models for the Next Generation of Software, O'Reilly Media, Inc.* Available at: http://www.oreilly.com/pub/a//web2/archive/what-is-web-20.html (Accessed: 1 March 2016).

OpenStreetMap Wiki contributors (2016) *Editing Standards and Conventions, OpenStreetMap Wiki.* Available at: http://wiki.openstreetmap.org/wiki/Editing_Standards_and_Conventions (Accessed: 3 June 2016).

OpenStreetMap Wiki contributors (2017) *Key:highway., OpenStreetMap Wiki.* Available at: http://wiki.openstreetmap.org/w/index.php?title=Key:highway&oldid=1332527 (Accessed: 30 January 2017).

Penn, A., Conroy, R., Dalton, N., Dekker, L., Mottram, C. and Turner, A. (1997) 'Intelligent Architecture: New Tools for the Three Dimensional Analysis of Space and Built Form', in *Proceedings of the First International Space Syntax Symposium.* London: The Bartlett School of Architecture, p. 30.1-19.

Penn, A., Hillier, B., Banister, D. and Xu, J. (1998) 'Configurational modelling of urban movement networks', *Environment and Planning B: Planning and Design,* 25(1), pp. 59–84. doi: 10.1068/b250059.

Peponis, J., Wineman, J., Bafna, S., Rashid, M. and Kim, S. H. (1998) 'On the generation of linear representations of spatial configuration', *Environment and Planning B: Planning and Design,* 25(4), pp. 559–576. doi: 10.1068/b250559.

Peponis, J., Wineman, J., Rashid, M., Kim, S. H. and Bafna, S. (1997) 'On the description of shape and spatial configuration inside buildings: convex partitions and their local properties', *Environment and Planning B,* 24, pp. 761–781.

Ratti, C. (2004) 'Space syntax:some inconsistencies', *Environment and Planning B: Planning and Design,* 31, pp. 487–499.

Sehra, S. S., Singh, J. and Rai, H. S. (2013) 'Assessment of OpenStreetMap Data - A Review', 76(16), pp. 17–20. doi: 10.5120/13331-0888 10.5120/13331-0888.

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C. and Haklay, M. (Muki) (2016) 'A review of volunteered geographic information quality assessment methods', *International Journal of Geographical Information Science,* 8816(August), pp. 1–29. doi: 10.1080/13658816.2016.1189556.

Serra, M., Hillier, B. and Karimi, K. (2015) 'Exploring countrywide spatial systems: Spatio-structural correlates at the regional and national scales', in Karimi, K., Vaughan, L., Sailer, K., Palaiologou, G., and Bolton, T. (eds) *Proceedings of the 10th International Space Syntax Symposium.* London: Space Syntax Laboratory, The Bartlett School of Architecture, University College London, p. 84.1-84.18.

Thomson, R. (2003) 'Bending the axial line: smoothly continuous road centre-line segments as a basis for road network analysis', in *Proceedings of the Fourth International Space Syntax Symposium.* London, p. 10. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.200.8956.

Turner, A. (2001) 'Angular Analysis', in *Proceedings of the Third International Space Syntax Symposium.* Atlanta, p. 30.1-30.11. Available at: http://discovery.ucl.ac.uk/35952/.

Turner, A. (2005) 'Could A Road-centre Line Be An Axial Line In Disguise?', *Proceedings 5th International Space Syntax Symposium,* pp. 145–159.

Turner, A. (2007) 'From axial to road-centre lines: a new representation for space syntax and a new model of route choice for transport network analysis', *Environment and Planning B: Planning and Design,* 34(3), pp. 539–555. doi: 10.1068/b32067.

Turner, A. (2009) 'Stitching Together the Fabric of Space and Society: An Investigation into the Linkage of the Local to Regional Continuum', in Daniel, K., Marcus, L., and Steen, J. (eds) *Proceedings of the Seventh International Space Syntax Symposium.* Stockholm: KTH Royal Institute of Technology, pp. 1–12. Available at: http://eprints.ucl.ac.uk/16184/ (Accessed: 24 June 2013).

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. doi: 10.1068/b31097.

Varoudis, T., Law, S., Karimi, K., Hillier, B. and Penn, A. (2013) 'Space Syntax Angular Betweenness Centrality Revisited', in Kim, Y. O., Park, H. T., and Seo, K. W. (eds) *Proceedings of the Ninth International Space Syntax Symposium.* Seoul: Sejong University, pp. 1–16.

Zielstra, D. and Zipf, A. (2010) 'A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany', in 13th AGILE *International Conference on Geographic Information Science.* Guimarães, Portugal, pp. 1–15. doi: 10.1119/1.1736005.