International Journal of Planning , Urban and Sustainable Development
ISSN 2311-9004

Evaluating Sustainable Housing Locations. Comparing the use of Network to Euclidean proximity in the scoring process

Abdul Naser Arafat and Mustafa Abed Al Musa – Birzeit University

Pedestrians tend to walk using the street network rather than walking on unpaved and undeveloped land. They usually follow straight line segments of the street network and the total path length following the street network is called the network distance. The estimation of the network distance depends on the characteristics of the network in terms of sidewalks and barriers.

Network distance has been recently introduced by researchers to replace the Euclidean distance in estimating proximity to services in land use models and in finding the suitable locations for sustainable housing projects. Despite the different research efforts and models that use network distance instead of Euclidean, Most suitability models are still using the Euclidean distance for estimating proximity to services such as transit, groceries and public schools.

This paper will compare the proximity scores to services obtained by network distance to the scores based on Euclidean distance. The study will show that there is a significant difference between the scores obtained by network distance and the scores obtained by Euclidean distance. The study will also show that the scores will be lower if the network distance is used. The study will also will take existing housing sites In Ramallah District- Palestine to show that using network distance is a better estimation especially for places that have connectivity issues such as cul-de-sacs. The study will point out the impacts of such difference on sustainable housing and derive recommendations for policy makers and housing organizations

Background:

The study of proximity depends largely of the magnitude and direction of travel distances. In this travel distance, people tend to follow straight line segments of the streets network. In other words, the travel distance is the distance traveled between origins to destinations following routes in a network.

Walking for example is one mode of the multi modal transportation systems and has its characteristics, when a student walks to school, he is doing a trip similar to cars and transit in other modes of the transportation system, the trip maker generally takes the shortest path to reach the school. This short path takes the student from this home to reach the school following within the minimum impendence, which is for walking mode, the walking time.

Following the research trends using GIS. Network distance is usually approximated by the Euclidean distance, which is a straight line distance or the absolute shortest distance between two points, or can be approximated by the Manhattan distance, which assumes that the distance between origins to the destinations follow a grid pattern. The state of practice used the straight line distance or the Euclidean distance in calculating the origin-destination distance. However, the Manhattan distance is rarely used in general but sometimes it is used for comparison purposes. Euclidean distance is commonly used in GIS for constructing surfaces using interpolation and statistical analysis. And that's why it is commonly used to represent the travel distance. This paper concentrates on GIS analysis that use the network distance to construct statistical surface to be used in preparing a proximity index layer. The resulting surface for a test area will be compared to the Euclidean distance commonly used. The paper will demonstrate that the resulting surfaces based on the network distance are different than the surface based on Euclidean or Manhattan distances.

Introduction:

Transportation research recently focus on the multi-modal transportation systems including interaction between all modes of transportation such as, driving, transit, biking and walking. The are four steps in the transportation modeling in a process called the four step modeling starting with the trip generation which is a function of the house hold characteristic and land use. The second step is the trip distribution where the trips are distributed between production zones and attraction zones. The third step is the model choice between driving, transit, biking and walking and the last step is the network assignment which concentrates on the routes and the network for the trips (Kutz, 2003).

Walking is an important mode of transportation. Modern urbanism and sustainable development urge the mixed used land uses to obtain walking, biking and transit friendly neighborhood that can lead to cleaner environment, less sprawl and a transportation system with less congestion (Levine, 2006). Furthermore, walking is important to all people because by walking we can reduce the danger of many deceases that began to spread because of using cars instead of walking.

Walking sometimes is not possible due to different factors like safety, distance, street characteristics and barriers and many other factors. This research studies walking to services (mainly schools) by the network distance approach and it aims to develop tools that can be useful in the study of the proximity to services for mixed used housing projects.

The study area for this paper is Ramallah district in Palestine. The paper will use schools to represent the services. The access to schools is estimated by Euclidean distance, Manhattan distance and network distance.

REVIEW OF LITERATURE:

In the absence of measured travel times to assess travel, the accuracy of topological accessibility to services estimations depend on the method used to estimate distance. Generally, distance measurement methods in land use research are one of three methods; Euclidean distance, rectilinear distance (Manhattan) and network distance. Many transportation studies involve measurements of distances between origins and destinations. In state of practice, researchers tend to use Euclidean distances rather than Network distances. This is because of the difficulties involved in calculating the network distances (El-Geneidy & Levinson, 2007).

The Euclidean distance is the straight line distance between two points using the coordinates of the two points; the distance can be calculated using the following equation:

$$D_{ij} = \sqrt{(X_j - X_i)^2 + (Y_j - Y_i)^2}$$

A third component can be also added to the equation if we want to include the z values. This distance type is a straight line distance, therefore can not consider the earth curvature or the street network (Wang, 2006). The same mathematical equation can be modified and generalized to approximate the street network as a grid pattern called the Manhattan distance. The Manhattan distance assumes a grid pattern that the measurement will follow. The modified equation that can be used for the Euclidean and the Manhattan distance can be written as following:

$$D_{ij} = \{(X_j - X_i)^\rho + (Y_j - Y_i)^\rho\}^{1/\rho}$$

For the Euclidean distance the value of $\rho = 2$ and for the Manhattan distance, $\rho = 1$ this means the Manhattan distance can be obtained by summing up the X and Y distance components which the easting and northing distances in the real world, this type of distance can give good result if the street grids follow a north-south, east-west pattern. If the street pattern is different, large computation error will be expected. The equation for the Euclidean distance is an equation for the circle while Manhattan distance equation is a straight line equation. And thus the value of ρ changes from the straight line to the circle portion (De Smith, 2007).

The network distance is the distance traveled by the person following a network route which is an accurate measure of the distance that depends on the road network characteristics and barriers. Researches historically assume that differences between Euclidean distance and network distance tend to be constant. This assumption is true only when variation in the network is minor and the person is not selecting his route by himself (self selection of routes). These self selection and variation in the network characteristic can lead to large differences in the distances computed by Euclidean, Manhattan and the network distance (EL-Geneidy & Levinson, 2007).

Calculating the network distance is more accurate than the Euclidean and Manhattan distance. However, the Manhattan distance is not an accurate measure for network distance but it can be considered as an efficient approximation when the variation in the network is low (Cromely & Mclafferty, 2002). In GIS there is no current function to calculate the less used Manhattan distance where this type of distance can be calculated using the X and Y coordinates for each point in the routes (Wang, 2006).

Network models are commonly used in GIS as road networks, telephone, power lines and water distribution. The network mainly consists of nodes and links and their characteristics. In transportation, regions are subdivided into traffic area zones (TAZ) in which its centroid is regarded as an access or egress node to the network. The roadways are considered the links and the junctions are considered the nodes. In a transportation network the flow is assumed in equilibrium and the people tend to follow the shortest route. This equilibrium status provides the mean for calculating the traffic volume for alternative routes as specified by Dijksta (1959). The calculation of the shortest route is done by summing up the travel time or distance for the successive links between an origin and a destination following different available routes. The minimum value calculated is considered the shortest route (Bolstad, 2002).

In most cases the network inside the TAZ is ignored and the calculation is done only for the network. The time of travel inside the TAZ is considered as access or egress time for entering or leaving the network node. Therefore, the estimation of distance can be easily obtained by the Euclidean distance. In GIS this distance can be calculated from any point to the network using the Nearest Euclidean distance with the neighborhood search radius (ESRI, 2017).

GIS can be used to calculate network and Euclidean distance. For the network distance the closest facility script in the network analyst for example, can be used to find the shortest route on the network (ESRI, 2017). Streets network with all the nodes, links, and characteristics can be defined and used as basis in the network analysis. GIS software (Arcgis 10.4) will find the shortest path beginning from the origin which is the incidents to the closest facility or destination. The result of the analysis will be a route layer containing all short path routes from the incident to the nearest facility on the network. (Bolstad, 2002)

Model builder, which is a graphic programming interface can be used to for complex analysis that needs many steps and difficult to be done on the fly. Model builder uses tools, input and output as geometrical shapes with arrows and links to construct a useful visual program. Complex programming can be done on model builder by making models which contain many ArcGIS functions and use this model as an object or a tool in another model (Carr & Zwick, 2007). Models can be also transferred to other programming languages such as Python script. Model builder can be considered as a graphical programming language that can do complex programs by connecting polygons and ellipses which represent the, input, output and tools. The model builder also have constrains on the looping processes. However, Iterations can be done on model builder and can include using the output as an input again in the loop. However, written scripts and object oriented programming languages are more flexible and the programmer can do more complex programs with complicated conditional structures. (ESRI, 2017).

Spatial Analysis can be used to construct surfaces and use these surfaces in other spatial analysis like spatial interpolation, stochastic modeling, map algebra and suitability ranking (Bolstad, 2002). Geo-statistical analyst and spatial analyst by ESRI can be used as extensions in ArcGIS for the spatial analysis. The user can use the tools on the fly or by constructing a model in the model builder using these tools; the user can also use these tools as objects in an object oriented programming language using ESRI geo-processing objects.

Many interpolation methods are presented in ESRI ArcGIS software. Stochastic and deterministic interpolation functions. Regression methods like kriging, inverse distance weighting and other type of spatial interpolation and smoothing functions. These functions use a regression function using points and direction and rotate this relation in all direction to get isotropic or anisotropic surfaces (Wang, 2006). These statistical and interpolation surfaces can be used as raster files for other types of map algebra and index suitability. Many mathematical functions can be used on raster files for the aim of combining multiplication, subtraction division, merging and appending, in addition to, reclassifying and ranking indexes. Model builder can also be used to construct complex models that uses map algebra and spatial reclassifying and ranking of suitability (Carr & Zwick, 2007)

METHODOLOGY AND PROCEDURES

Network distance can be obtained from a network property approach by measuring the length of street segments as a percentage of the whole street network (O'nell, 2004), or by measuring the actual distance travelled (Zhao et al, 2003). The use of travel time may be more sophisticated and take additional variables into consideration. However, in measuring network distance barriers can be included to give a more accurate indication of travel distance. Arafat, Steiner and Bejleri (2008) compared network distance to Euclidean and Manhattan distance in research on school siting. Their research found that the use of network distance gives a better estimation for walking distance than Euclidean or Manhattan distances. Additionally they found that the catchment area for population, which is an accessibility indicator, is exaggerated when using a Euclidean buffer.

The network distance is used in transportation research to build accessibility indices in Texas (Bhat et al, 2002), where, the travel distance had been obtained from travel surveys which may not be available on a disaggregate level. An alternative methodology can be used to generate the network distance at a parcel level using ArcGIS network analyst which is software that can calculate distance from origins to destination following the road networks. In this methodology, the shortest network distance can be measured from each origin to each destination (Arafat et al, 2008).Traditionally, proximity as straight line distance has been used as an accessibility measurement in deterministic land-use models (e.g. see, Zwick & Carr, 2007) while gravity models have been used in statistical and stochastic models (e.g. see, Waddell, 2002). There are also other measurements of accessibility used in statistical models such as opportunity access (e.g. see, Handy, 2004; Hanson 2004).

However, using gravity models on a parcel-level scale requires generating huge origin destination matrices that contain billions of records for large study areas. These origin destination matrices exceed the capacity of hardware and software used in the analysis. Therefore, a methodology to create smaller representative datasets can be used. The distance estimation component can be performed by network analysis. In this paper GIS network analysis will be done to find the shortest path from each residential parcel to services. To do that, a network of all the roads in the county is to be constructed using ArcGIS software. A model to find the shortest network distance from each parcel to the services is prepared for that purpose (ESRI, 2017). The values for distance will be used as point values for the residential parcel points where statistical analysis will be done using ArcGIS geo-statistical wizard and spatial analyst to construct surfaces that can be used as proximity surfaces.

GIS Analysis

Creating distance a raster is one of the simplest models to measure accessibility. This is traditionally created using the Euclidean distance raster tool provided by the ArcGIS software. Creating raster grids using other type of distances such as the Manhattan and network distance is not provided by the ArcGIS software. This paper uses Python programming and other ArcGIS tools programmatically to create distance raster grids based on Manhattan and network distance. The Manhattan distance grid is created by map algebra operations that are performed on the output of ArcGIS Euclidean rasters. The procedure is performed in Python and produced as a customized tool for Manhattan distance. The paper also uses a customized network distance raster which provides the shortest network distance to the nearest facility. The network distance raster is also customized by Python and estimates distances from all the locations to their nearest destinations and generates a network distance raster instead of a Euclidean distance raster.

To reduce the memory used by the tool and to increase the processing speed, the research used random origins and uses them to estimate the distance from origins to destinations using Network analyst. The random origins are generated by the user before using the tool. The use of the random sample is optional and the user can replace them by the whole population. The output raster is created by interpolating the results of the distance estimated for each origin point to the closest destinations.

Spatial and geo-statistical analysis is used to construct, combine and compare surfaces that where obtained from point feature classes (ESRI, 2017). Point data will be interpolated using different techniques. The IDW surface is used in a further study using map algebra for raster files to divide the IDW Euclidean distance surface with a network distance surface to have an idea about the network circuity which is the ratio of the Euclidean distance to the network distance.

Statistical analysis

Statistical Package for Social Sciences SPSS will be use to describe and compare the distances. Using ArcGIS zonal statistics the distances where summarized to the built areas which where compared by SPSS for significant differences

ANALYSIS AND RESULTS

The distance to the nearest school is estimated by the Euclidean distance tool in ArcGIS. Figure 1 shows the Euclidean distance raster which estimates the Euclidean distance from any point in the district to nearest school.

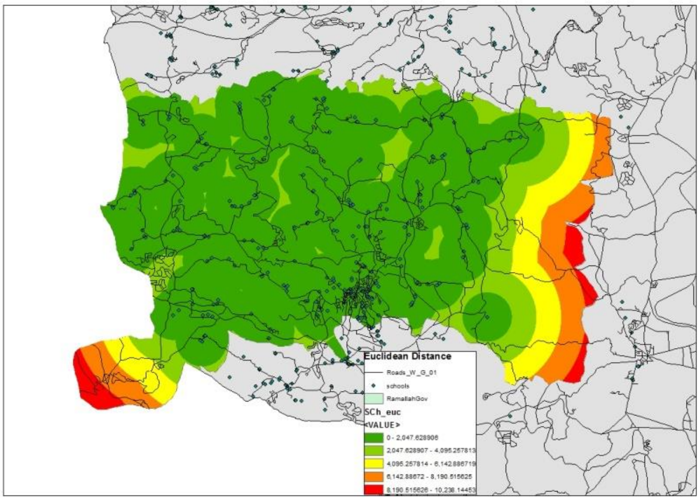

Figure 1: Euclidean distance to the nearest school

The Manhattan distance is also estimated using a customized Manhattan distance using programing in the GIS environment, the following figure shows the rectilinear Manhattan distance from any point in the district to the nearest school.

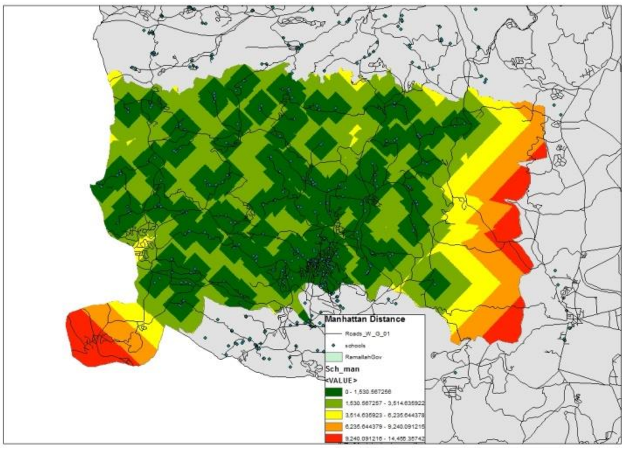

Figure 2: Manhattan distance to the nearest school

The network distance is also estimated using a customized network distance using programing in the GIS environment, the following figure shows the network distance from any point in the district to the nearest school.

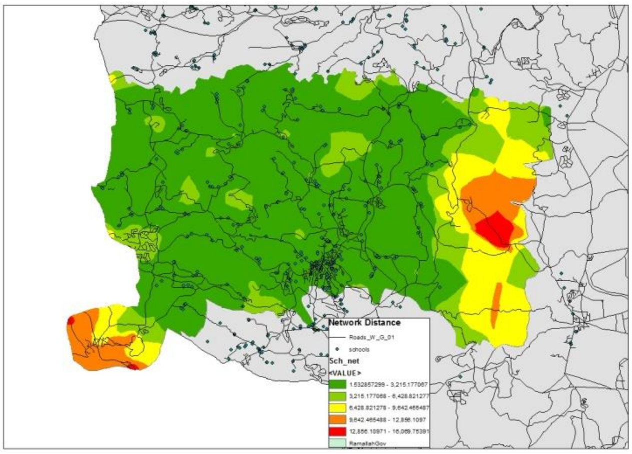

Figure 3: Network distance to the nearest school

Using SPSS, a paired sample comparison is performed between the three distances. The following table shows the result of the comparison

Table 1 comparing network and Manhattan distances to Euclidean distance

Pair	Distances	Paired Differences		Sig. (2-tailed)
		Mean	Std. Deviation	
Pair 1	Euclidean - Manhattan	-271.625387	346.8526887	0
Pair 2	Euclidean - Network	-730.057123	1100.090369	0

The table shows clearly that mean Manhattan distance is 271.62 meters more than the Euclidean distance. It also shows that the mean network distance is 730 m more than the

Euclidean distance. The differences are statistically significant on 0.95 level. This means that Euclidean distance underestimates the real distance to the nearest school. The distances are also summarized for each built area in the district. Table 2 shows the results.

Table 2 Average distances for built areas to nearest schools

Type	Euclidean	Manhattan	Network
Camp	466.26325	642.75875	555.64775
City	362.213	462.3865	661.7115
Town	508.82	648.06025	838.25875
Village	594.1632986	754.269809	1001.842486
Settlement	1913.491324	2442.01835	3424.149757
Total	986.7109264	1258.33631	1716.76805

From the table we can see that the network distance is more than the Manhattan or the Euclidean distance in each of the built up areas. Manhattan distance represent the best connectivity in the ideal street network. Having larger distances than the Manhattan distance points to the fact that there is connectivity problems in the road network. Which indicates that a person needs to walk for larger distances to reach a destination using the street network. One of the walkability indices Pediatrician route directness ratio PDR is defined as the network distance divided by the Euclidean distance. The perfect ratio is 1-1.41 while more than 1.41 values means problems of walkability in street networks. The following map shows the PDR values for each built-up area.

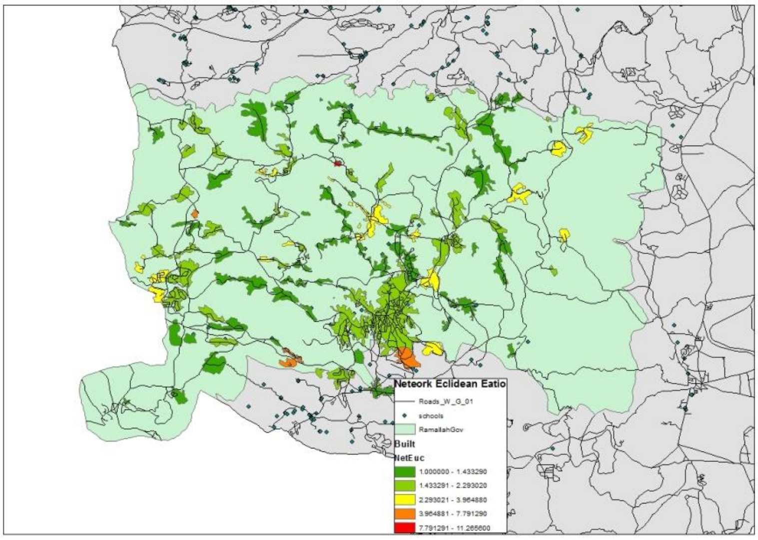

Figure 4: PDR ratios for buildup areas

While all the Manhattan to Euclidean distance PDR are between 1- 1.41 we can see than many built-up areas has a PDR of more than 1.41 which is due to the lack of roods in rural areas or road obstacle such as military checkpoint in urban areas. The following chart compares the distances according to the built up type

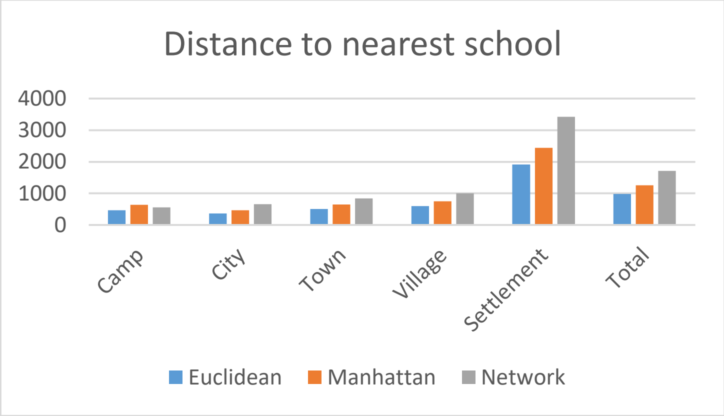

Figure 5 comparing distances according to built up area type

From the charts, the distance increase for the rural areas such as the towns and villages and it's decreasing for urban areas which include cities and camps near or inside cities. The chart also shows the trend of network distance more than euclidean and manhattan distances. As a collateral result the charts shows how much the israeli settlements are close to Palestinian schools which represent fears and uncomfortable conditions for students going to schools

DISCUSSION AND CONCLUSIONS

We can see from the charts and analysis that the network distance is more than the Manhattan and the Euclidean distances. Considering proximity to service by walking, the 5 and 10 minutes of walking time can be translated to 400 m and 800 m respectively. In land use and sustainable development planning, planners use mostly Euclidean distance, this means that the 800 meter might be doubled if we follow the street network. People might be willing to walking 800 m. However, many of them might not be willing to walk 1600 m. The results of this research shows clearly that Euclidean distance underestimate the actual distance.

The research result also shows that rural areas have a PDR ratio more than urban areas which suggest the street connectivity is better in urban areas. The connectivity of streets is also better if the network follows the grid system as in Manhattan grids. The connectivity will be perfect also when the origin and destination are on the same street. In this case the PDR ratio equals 1.

In suitability models (Zwick, 2007), the land use proximity uses Euclidean distance. This paper recommends that the use of network distance is better estimation for distance and therefore should replace the Euclidean distance in the proximity for incentives and disincentives.

Housing agencies usually evaluate proximity by Euclidean distance. This research suggest that the use of network estimation of impedance should be used instead.

LIMITATIONS AND FURTHER RESEARCH

1. The use of network distance should be limited to transportation modes of walking, biking, transit and driving. Water and air transportation are not within the scope of this research.

2. The Euclidean distance should be used for incentives and disincentives that does not follow the street network such as the proximity to open water, pollution and noise generators

3. The study was controlled by certain hardware and software limitations that can be summarized by the weak interface between the network analyst and the scripting language because the network analyst is designed to work on the fly or on the model builder. In both cases the computer works on interactive sessions which were not the case in the study. This none interactive environment made it difficult to use certain function that was important in reducing running time and increasing the efficiency of the developed tools.

4. The study was done assuming no barriers on the street network and can be modified by defining the barriers on the network to obtain more accurate result in the future.

5. More statistical analysis on parcel zonal data can be analyzed using the network distance surface as a walkability index surface.

References:

Bolstad, P. (2002). GIS Fundamentals: A First Text on Geographic Information Systems, Eider Press, White bear Lake, Minnisota.

Carr M. and P. Dwick (2007). Smart Land-Use Analysis: The LUCIS Model, ESRI press, Redlands, California.

De Smith, G. et al (2007). Geospatial Analysis: a comprehensive guide. Extended Euclidean and Lp-metric distances.

Dijkstra, E.W. (1959): A note on Two Problems in Connexion with Graphs, Numerische mathematik, 1, 269-271.

El-Geneidy, A. and D. Levinson (2007) "Network Circuity and the Journey to Work," Presented at the University Transport Study Group Conference at Harrowgate, England, January 3-5 2007.

Environmental Systems Research Institute (2017). In Spatial analyst web help. Retrieved February 15, 2017 from
http://desktop.arcgis.com/en/arcmap/10.4/analyze/main/what-is-geoprocessing.htm

Haywood, Ian et al (1998): An Introduction to Geographical Information systems, Addison Wesley Longman Limited.

Kutz, M. (2003), Handbook of Transportation: Transportation Planning and Modeling, McGraw Hill.

Lee, R.W., and R. Cervero. 2007. The effect of housing near transit stations on vehicle trip rates and transit trip generation: A summary review of available evidence. [Cited 15 April 2009]. Availble from http://www.reconnectingamerica.org/assets/Uploads/hcd_tod_resource_paper_2008_09_20.pdf.

Moshe Sniedovich (2007). Dijkstra's Algorithm revisited, "the OR/MS Connexion", Department of Mathematics and Statistics The University of Melbourne Parkville, VIC 3052, Australia

S. Biba , K. M. Curtin & G. Manca (2010) A new method for determining the population with walking access to transit, International Journal of Geographical Information Science, 24:3, 347-364 To link to this article: http://dx.doi.org/10.1080/13658810802646679

Srour, I., Kockelman, K. & Dunn, T., 2002. 'Accessibility indices: a connection to residential land prices and location choices'. Paper presented at the 81st Annual Meeting of the Transportation Research Board, January 13-17, 2002: Washington, DC.

Wang, Fahui (2006) Geographic information systems: Mathematical models, CRC Press.