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.

Network Circuity and the Location of Home and Work

David Levinson
Department of Civil Engineering
University of Minnesota
500 Pillsbury Drive SE
Minneapolis, MN, 55455 USA
dlevinson@umn.edu
+01-612-625-6354

&

Ahmed El-Geneidy
Department of Civil Engineering
University of Minnesota
500 Pillsbury Drive SE
Minneapolis, MN, 55455 USA
geneidy@umn.edu
+01-612-624-8282

Abstract

In an urban context people travel between places of residence and work destinations via transportation networks. Transportation studies that involve measurements of distances between residence and work locations tend to use Euclidean distances rather than Network distances. This is due to the historic difficulty in calculating network distances and based on assumptions 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 when self-selection is not present. In this paper we use circuity, the ratio of network to Euclidean distance, as a tool to better understand the choice of residential location relative to work. This is done using two methods of defining origins and destinations in the Twin Cities metropolitan region. The first method of selection is based on actual choice of residence and work locations. The second is based on a randomly selected dataset of origins and destinations in the same region. The findings of the study show circuity measured through randomly selected origins and destinations differ from circuity measured from actual origins and destinations. Workers tend to reside in areas where the circuity is lower, applying intelligence to their location decisions. We posit this because locators wish to achieve the largest residential lot at the shortest commute time. This finding reveals an important issue related to resident choice and location theory and how resident workers tend to locate in an urban context.

Keywords

Network structure, travel behavior, transport geography, commuting, circuity

Introduction

"Man walks in a straight line because he has a goal and knows where he is going; he made up his mind to reach some particular place and he goes straight to it" (Le Corbusier, 1929). The presence of a transportation system (including networks and modes) dissuades people from achieving traveling to their destinations in a straight line by providing the opportunity to move faster if more circuitously. The straight line distance between two points in an urban environment is known as the Euclidean distance. Meanwhile the distance that a person travels along a transportation network between origins and destinations is known as network distance. The network distance is affected both by travelers' choices and the spatial distribution of the network. Selection of any random pair of points in an urban environment and measuring circuity (the ratio of network to Euclidean distances) leads to a different answer than the actual selection of an origin and destination by locator-travelers.

Distances in transportation research can be measured using geographic information systems (GIS), which enable calculation of distances between origins and destinations in three forms: Euclidean distance, network distance, and Manhattan distance. Manhattan distance is not commonly used in transportation research since it is generally meaningful on a grid system, which is true only in certain urban contexts. Euclidean distance is the airline distance measured between origins and destinations, while the network distance, which is a more realistic representation of movements between origins and destinations, is the distance between origins and destinations measured along a transportation network. Figure 1 shows differences between network distance and Euclidean distance.

Insert Figure 1

Historically, many researchers used either the Euclidean distance or the Manhattan distance, which both rely on simple mathematical calculations (Pythagorean Theorem), in calculating distances in transportation research due to the complexity of the process computing an accurate network distance measure. Network distances can be computed using a number of techniques, one way of computing shortest path relies on the theory developed by W. R. Hamilton (1805-1865) and known as the *Hamiltonian cycles* (Miller & Shaw, 2001).

Circuity has been examined by a number of researchers in a variety of contexts. Newell (1980) indicated that Network distance measured for a randomly selected set of points in an urban environment is about 1.2 times the Euclidean distance, though this is true only for a certain types of networks. Other research (O'Sullivan & Morral, 1996) finds circuity factors of 1.21 – 1.23 at various transit station catchment areas. The measure has also been used at the national level (Ballou *et al.*, 2002), and for pedestrian and bike travel (dubbed pedestrian route directness) (Dill, 2004), with much higher values than is observed for automobile travel. The measure has also been considered by Axhausen et al. (2004) using GPS traces of actual travelers route selections, finding that many actual routes experience much higher circuity than might be expected.

Transportation studies that involve measurements of distances between residence and work locations have tended to use Euclidean distances rather than Network distances. This is due to the historic difficulty of calculating network distances and based on assumptions that differences between Euclidean distance and network distance are small and constant. This assumption only holds when variation in the network is minor and when self-selection is not present. Residents choose homes weighting the attributes of accessibility to work, shopping, schools, quality of neighborhood life, availability of public service, amenities related to his housing choice (number of bedrooms, bathrooms, etc.) and costs of living in such area (McFadden, 1977). Since work is one of the most critical aspects of this decision process, in this research we try to understand the effects of choice of residential location relative to actual work location.

In contrast to some research (Guiliano & Small, 1993), though consistent with the traditional models of urban economics, we posit that journey to work importantly affects location choice, and propose to test that by examining the relationship of network to Euclidean distance.

The standard model of urban economics depends on the basic assumption that choice of residential location is based on tradeoff between commuting cost and land cost (Mills, 1972). In the classic monocentric urban economic model, house location relative to work is identical to house location relative to the center of the city, and thus how much land costs. The cost of land tends to decrease with the increase in Euclidean distance from the center, while keeping other factors affecting land value constant at their mean values. Network distance, on the other hand, is an indicator of how much travel actually takes place (and is more closely related to travel time), which has implications for congestion, pollution, and travel behavior and activity patterns. The commute time tends to increase with the increase in the network distance, while keeping all other factors affecting travel time constant at their mean values.

Cities of course are not monocentric, yet we hypothesize Euclidean distance between home and work is still correlated with land and housing costs, and that individuals will choose greater Euclidean distances to improve the quality and size of their home. Individuals who minimize circuity lie on the frontier with the maximum house/lot at the minimum travel time (as the network distance is closer to a straight line).

In this paper we measure network circuity using two methods of defining origins and destinations in the Twin Cities metropolitan region. The first method of selection is based on actual choices of residence and work locations. The second is based on a randomly selected datasets of origins and destinations in the same region. The findings of this research assist in better understanding choice of residential location relative to their work. We believe this is the first paper to compare Euclidean distance to network distance while taking into account self-selection as a factor.

The paper next details the research design of the paper. This is followed by a description of the data sets used herein. It next constructs a model to predict the network circuity as a function of network structure and transportation geography. The models are analyzed, and finally conclusions are drawn.

Research Design

The relationship between network distance and Euclidean distance is important in several contexts. Euclidean distance indicates how much area is passed between two points. We posit that individuals would like to have the most space available at the least travel and monetary cost. This implies that, all else equal, residence and work locations will be chosen where the network circuity is at a minimum. We expect to find that circuity is lower for actual home and work pairs than for the random set of home and work pairs that have been used in previous research (Newell, 1980), as people can select how to arrange their activities on the network.

To test this hypothesis we propose to compare network and Euclidean distance using several sets of origins and destinations. The first is a set of origins and destinations defined based on residence and work locations of workers in the Twin Cities metropolitan region. The second dataset uses the same origins and destinations, but randomizes the matching of origins and destinations, and so is most analogous to previous analyses of network circuity, as reported in Newell (1980). This is done in a stepwise method to capture the effects of various distances between random origins and destinations. The third dataset uses the same origins and destinations, but randomizes the matching while trying to ensure the same statistical distribution of network distance (but allowing the Euclidean distance to vary). The fourth uses the same origins and destinations but randomizes the matching while retaining the same statistical distribution of Euclidean distance, while allowing the network distances to vary.

Insert Table 1

We expect that circuity in Case 1 is the lowest of the four cases, as that case represents intention on the part of travelers. In addition the research design will predict the network distance of OD pairs as a function of Euclidean distance, while controlling for location of origin and destination and interaction of the two, and type of network present.

Data

The Longitudinal Employer-Household Dynamics dataset (LEHD) used here is a comprehensive dataset that includes people's place of residence identified at the Census Block level of analysis and their employment location identified at the same level. This origin and destination dataset can be used to test our research question. In order to decrease the complexity of calculations a random sample of resident workers is selected (5,000 observations) from the LEHD dataset. The LEHD data set for the year 2002 contains 1,422,980 observations aggregated at the census block level of analysis. This home-work pair matrix contains records with multiple workers sharing both origin and destination block. Duplicate records were added to the data to generate a new data set with 2,377,157 actual home-work pairs. From the 2,377,157 a sample of 5,000 home-work pairs was selected to be used in the calculation of case 1. Both Euclidean and network distances are calculated for the 5,000 home-work pairs. Figure 3 shows the distribution of the selected home and work locations in the Twin Cities region.

To prepare data for cases 2, 3 and 4, two new samples are generated. The first includes 200 randomly distributed points in the region, while the second includes 1,000 randomly distributed points. The first sample is used as origins and the second is used as destinations to generate a random origin destination (OD) matrix. The sample included 200,000 OD pairs. Both Euclidean and network distances are calculated for the entire random matrix. In order to generate data for cases 3 and 4, Euclidean and network distances are rounded to the nearest 10 meters in both matrices (random and home-work). From the random 200,000 OD pair matrix a sample of 5,000 observations is selected that have the same Euclidean distances as the observed sample. This sample is used to generate the data needed for case 3. Similarly a sample of 5,000 OD pairs is selected from the random 200,000 OD pair matrix, while fixing the network distance to generate data for case 4. Meanwhile for case 2 a random sample of 5,000 observations is selected from the 200,000. The average Euclidean and network distance in each case is displayed in Table 2. Distances are measured in meters.

Euclidean distance is calculated from the coordinates of the centroid of the census block of origin and destination. Network distance is calculated using the ArcGIS module Network Analyst on a road network provided by MetroGIS. An implicit assumption is that travel takes place on the road network, though we do not know the actual mode or path used for travel in the LEHD database. Accordingly we use the shortest network distance between two points as a proxy for the actual network distance.

Insert Figure 2 and Table 2

It is clear from observing Table 2 that differences exists in term of the distribution between the randomly selected sample (case 2) and the distribution of all the other samples. Accordingly comparing case 2 directly with the other cases is inaccurate. Case 2 is divided into various subsamples that will be selected randomly from the 200,000 sample for the sake of understanding the difference between randomly selected pairs and home-work pairs based on the Euclidean distances.

Analysis

Using network distance as the dependant variable and Euclidean distance as the independent variable, while assigning the value of zero to the intercept, regression models are estimated for each case. Table 3 includes the output of the regression analyses showing the network circuity and the average network and Euclidean distances. All models had a sample size of 5,000 observations. The results of the regression differ from the results in Table 2 because of different averaging techniques: Table 2 averages all observations and then takes the ratio, while Table 3 essentially takes the ratio for each individual and averages that.

The analysis shows that differences between Euclidean and network distances measured through randomly selected origins and destinations tend to differ from distances measured based on actual origins and destinations that are based on selection of residence and work place. It is important to note that the pairs derived based on the home-work relationship has a lower average circuity (1.18) compared to all the other randomly selected points (1.22 for case 3, 1.25 for case 4). This observation holds even when matching the Euclidean or network distances. In other words, worker-locators tend to choose jobs and residences where the circuity is lower, applying intelligence to their location decisions. This finding reveals an important issue related to resident choice and location theory and how resident workers tend to locate in an urban context: the efficiency of the network cannot be assessed independently of how travelers use it. Using a t-test it was demonstrated that the home-work relation tend to be statistically different from all the other randomly observed measures of circuity.

Observing case 2, the circuity decreases with the increase of the both Euclidean and Network distances. This relationship is represented in Figure 3.

Insert Table 3 and Figure 3

While the model in Table 3 shows the general trends in the relation, this method of analysis does not account for the current network structure. A linear regression model is used to determine spatial and network factors affecting the circuity as the dependant variable. The sample used included 5,000 observed home-work pairs in addition to a 5,000 randomly selected pairs.

Four different urban contexts (rings) are identified in the Twin Cities metropolitan region, each with a different type of street network: the urban ring, which includes cities of Minneapolis and Saint Paul, has the most grid-like network, while the suburban rings have network topologies which are more tree-like. The still rural outer ring also has a grid-like network, but the grid is at a much larger scale than in the center city.

Insert Figure 4

A 2 kilometer buffer is generated around the line representing the Euclidean distance. This buffer is used to calculate the characteristics of the network in the area between origins and destinations. Characteristics include the number of nodes generated due to intersections between streets, streets and freeways, and/or freeways with freeways. The length of freeways and streets in the buffer are also included in the characteristics section. The shortest network distance between each origin and destination is included to account for the length of the trip. Each pair of origin and destination intersects with one or more rings. A set of dummy variables representing which region the line representing the Euclidean distance intersects is included. The final functional form of the model will show as follows:

Circuity = f (Street-street nodes, street-freeway nodes, freeway-freeway nodes, Freeway length, street length, shortest network distance, ring dummy, home-work dummy).

The home work dummy is added to measure the effect of home location relative to work on circuity. Table 4 includes the output of the model.

Insert Table 4

Observing Table 4 it is clear that if the line measuring the Euclidean distance crosses rings 1 or 4 a negative effect is present, while crossing rings 2 or 3 a positive effect is present. (Some buffers intersect multiple rings, allowing us to avoid correlation problems). This observation indicates the presence of a unique network structure in each ring. Ring 1, which represents the core cities of Minneapolis and Saint Paul, has a well connected grid system, as does the rural ring 4 (though at a lower density). Rings 2 and 3 have more tree-like, and less direct, suburban road networks.

Surprisingly, the number of street-street nodes and freeway-freeway nodes has a statistically significant and positive effect on circuity. Meanwhile the number of nodes representing intersections between the freeway system and the street system has a statistically insignificant (though negative) effect on circuity.

Both freeway length and network length have a statistically significant negative effect on circuity, so the more roads, the more direct the path possible, which is expected. The actual network distance, which is included as a control variable to represent scale, shows a statistically significant positive effect on circuity, which is an expected effect.

The home-work dummy variable (the policy variable) did show a statistically significant negative effect on circuity.

Conclusions

In this paper we use the ratio of network to Euclidean distance (network circuity) as a tool to better understand the relation between choice of residential location relative to work. This is done using two methods of defining origins and destinations in the Twin Cities metropolitan region. The first method of selection is based on actual choice of residence and work locations. The second is based on a randomly selected dataset of origins and destinations in the same region. The findings of the study show that circuity measured through randomly selected origins and destinations differ from circuity measured from actual origins and destinations. Workers tend to choose commutes with lower circuity, applying intelligence to their location decisions. We posit this is because locators wish to achieve the largest residential lot for the shortest commute time.

The results indicate that people are selecting network circuity ratios that are 0.056 smaller than random, while controlling all other factors affecting the selection of home locations. While that number may not sound large, this is 0.056 better than the 1.22 (case 3) random result, which is about 25% better (since the best possible ratio would be 1.0). Given all of the other constraints individuals face when finding housing and jobs in a multi-worker context and on inefficient networks, we conclude that maximizing land while minimizing commute remains an important factor in urban location decisions. An important corollary of these findings is that the efficiency of the network cannot be assessed independently of how travelers use it. The circuity that users face depends on what they do.

References

Axhausen, K., Schonfelder, S., Wolf, J., Oliveria, M., & Samaga, U. (2004). *Eighty weeks of gps traces, approaches to enriching trip information.* Paper presented at the Transportation Research Board Annual Meeting.

Ballou, R., Rahardja, H., & Sakai, N. (2002). Selected country circuity factors for road travel distance estimation. *Transportation Research Part A, 36*, 843-848.

Dill, J. (2004). *Measuring network connectivity for bicycling and walking.* Paper presented at the Transportation Research Board Annual Meeting.

Guiliano, G., & Small, K. (1993). Is the journey to work explained by the urban structure? *Urban Studies, 30*(9), 1485-1501.

Le Corbusier. (1929). *The city of to-morrow and its planning* (P. Clarke, Trans. Eight ed.). New York: Dover Publications, Inc.

McFadden, D. (1977). *Modeling the choice of residential location* (No. 477). New Haven, Connecticut: Cowels Foundation for Research in Economics.

Miller, H., & Shaw, S. (2001). *Geographic information systems for transportation: Principles and applications*. New York: Oxford University Press.

Mills, E. S. (1972). *Studies in the structure of the urban economy*. Baltimore: Johns Hopkins Press.

Newell, G. (1980). *Traffic flow on transportation networks.* Cambridge Massachusetts: MIT Press.

O'Sullivan, S., & Morral, J. (1996). Walking distances to and from light-rail stations. *Transportation Research Record*(1537), 19-26.

Table 1: Four Cases Analyzed

Euclidean Distances Network Distances	Fixed to Observation	Random
Fixed to Observation	Case 1: (Observed)	Case 4
Random	Case 3	Case 2: (Literature)

Table 2: Average Distances

Case	Relationship	Average Network Distance	Average Euclidean Distance	Ratio of Averages
1	Home-work	17,845	14,746	1.21
2	Random	54,891	45,926	1.20
3	Euclidean Distance Matched	19,473	14,987	1.30
4	Network Distance Matched	18,134	14,357	1.26

Table 3: Network Distance = f(Euclidean Distance) Models

Case	Description	*Circuity Coeff.*	R^2	Avg. Net. Dist.	Avg. Euc. Dist.
1	Home-work	1.18	0.99	17845	14746
3	Random with Network distribution equal to the HW distribution	1.22	0.99	18134	14357
4	Random with Euclidean distribution equal to the HW distribution	1.25	0.98	19473	14987
2-1	Euclidean distance less than 5,000 meters	1.58	0.85	5250	3295
2-2	Euclidean distance less than 10,000 meters and greater than 5,000 meters	1.42	0.94	11021	7731
2-3	Euclidean distance less than 15,000 meters and greater than 10,000 meters	1.34	0.97	16986	12639
2-4	Euclidean distance less than 20,000 meters and greater than 15,000 meters	1.30	0.98	22845	17549
2-5	Euclidean distance less than 25,000 meters and greater than 20,000 meters	1.27	0.98	28660	22558
2-6	Euclidean distance less than 30,000 meters and greater than 25,000 meters	1.25	0.99	34376	27539
2-7	Euclidean distance less than 35,000 meters and greater than 30,000 meters	1.23	0.99	40072	32536
2-8	Euclidean distance less than 40,000 meters and greater than 35,000 meters	1.22	0.99	45762	37554
2-9	Euclidean distance less than 45,000 meters and greater than 40,000 meters	1.21	0.99	51267	42519
2-10	Euclidean distance less than 50,000 meters and greater than 45,000 meters	1.20	0.99	56745	47457

Table 4: Circuity = f(Network Attributes) Models

	Coefficients	*t Stat*
Intercept	1.4273	161.64
Number of street-street nodes	0.0002	9.50
Number of street–freeway nodes	-0.0001	-1.15
Number of freeway-freeway nodes	0.0006	8.91
Freeway length	-0.0013	-10.34
Street Length	-0.0010	-11.45
Network Distance	0.0061	13.51
Dummy if buffer intersect Ring 1	-0.0096	-1.22
Dummy if buffer intersect Ring 2	0.0136	1.74
Dummy if buffer intersect Ring 3	0.0097	1.49
Dummy if buffer intersect Ring 4	-0.0508	-6.78
Home Work Dummy	-0.0568	-8.99

$R^2 = 0.11$ N= 10,000

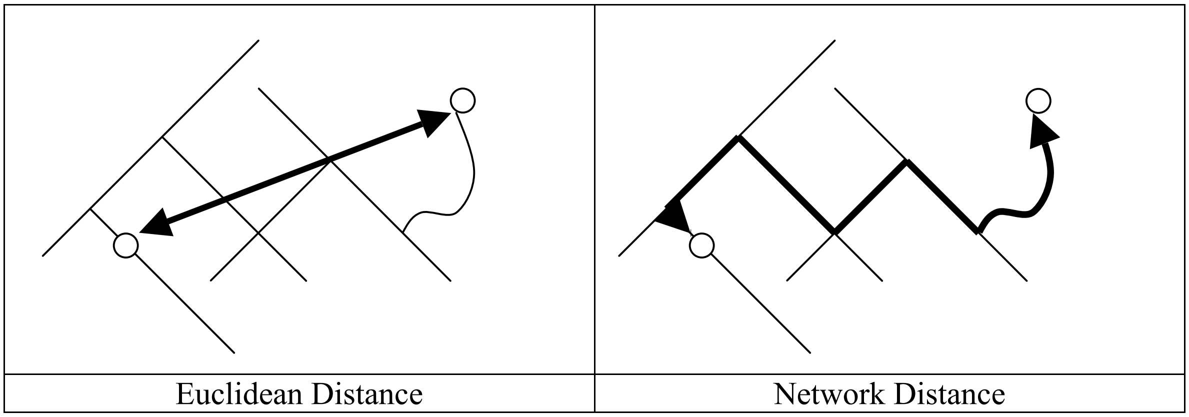

Figure 1: Difference Between Euclidean and Network Distances

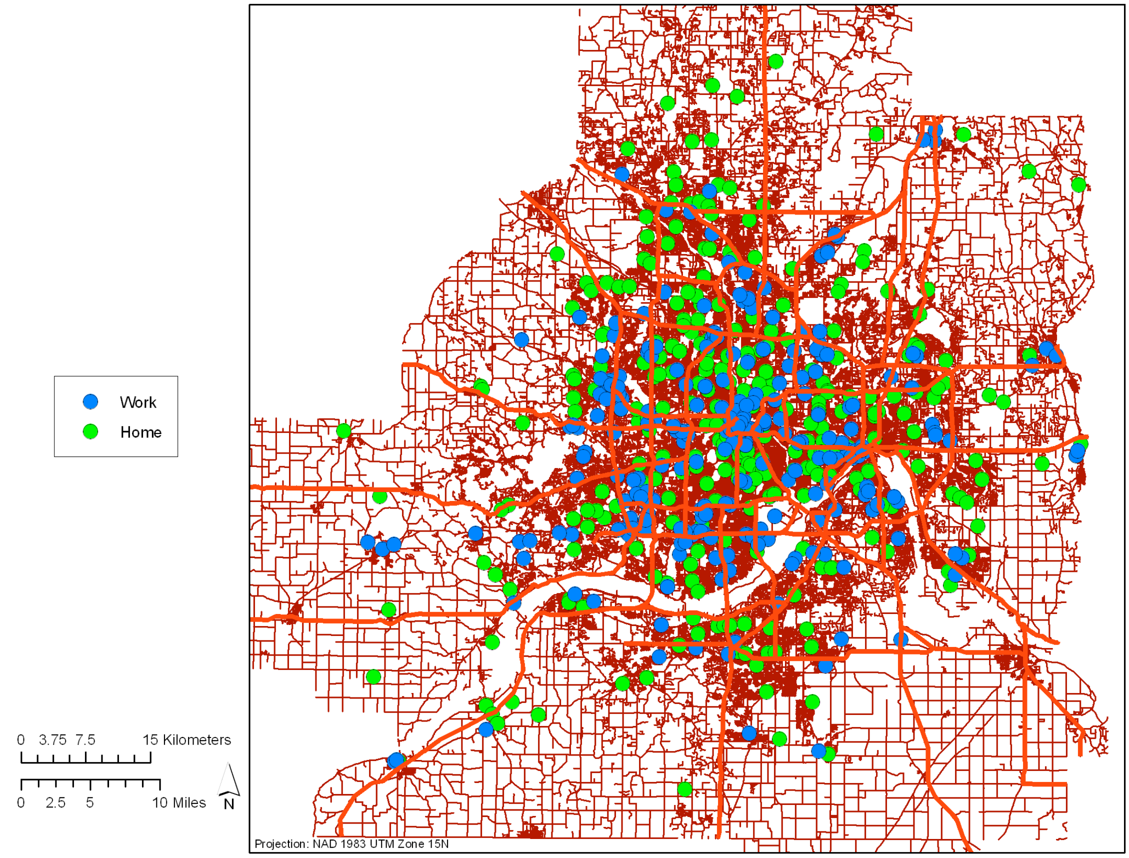

Figure 2: Distribution of home and work locations in the Twin Cities Region

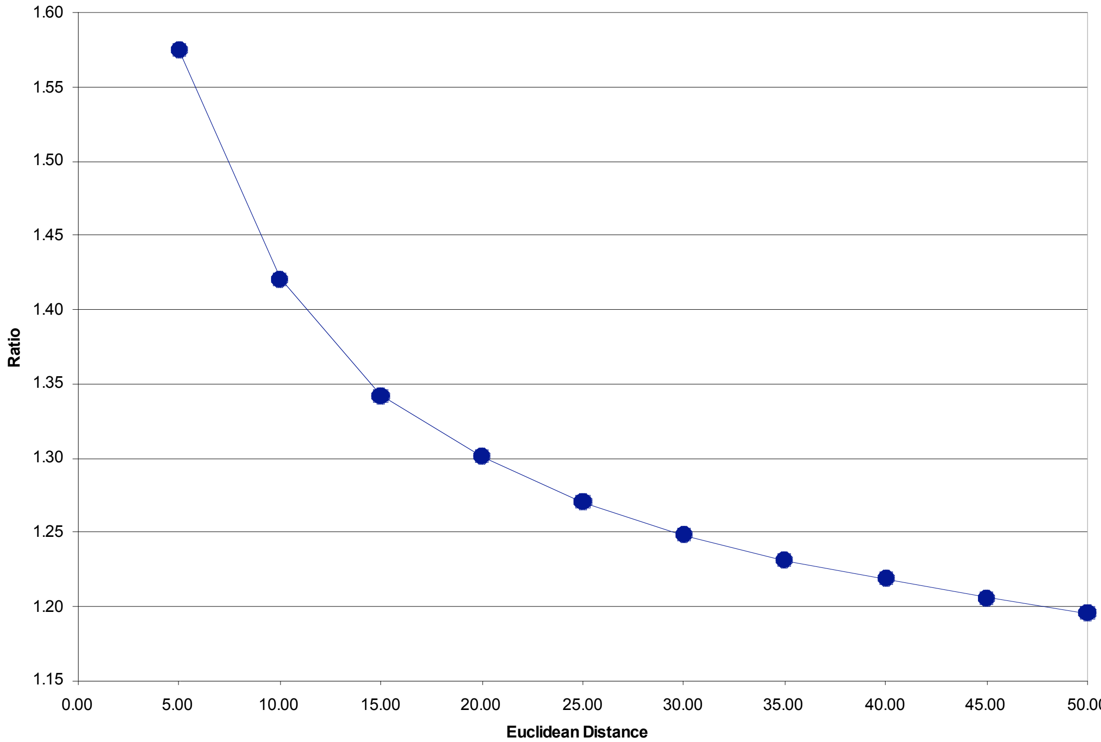

Figure 3: Circuity (the ratio between Network and Euclidean distance) using random sample stratified by Euclidean distance

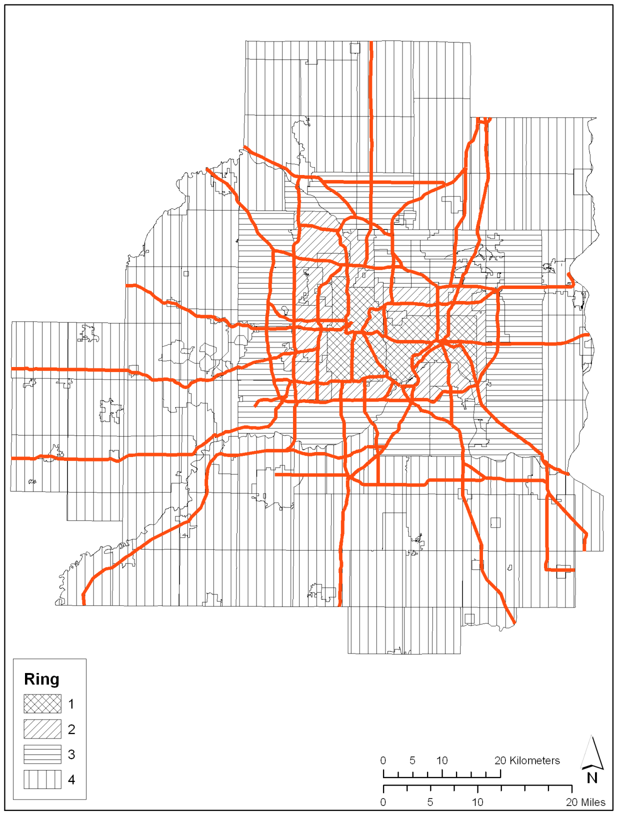

Figure 4: Study Zones