Associations between Road Network Structure and Pedestrian-Bicyclist Accidents

Yuanyuan ZHANG (corresponding author)
PhD Student, School of Transportation Engineering, Tongji University
Visiting Scholar, Safe Transportation Research & Education Center
Institute of Transportation Studies, UC Berkeley
2614 Dwight Way, Mail code #7374
Berkeley, CA 94720-7374
Tel: 315 706 6231
Email: yuanyuanzhang@berkeley.edu;
chanel830719@hotmail.com

John BIGHAM
GIS Program Manager
Safe Transportation Research & Education Center
Institute of Transportation Studies, UC Berkeley
2614 Dwight Way, Mail code #7374
Berkeley, CA 94720-7374
Email: jbigham@berkeley.edu

Zhibin LI
PhD student, Southeast University, P. R. of China
Visiting Scholar in Safe Transportation Research & Education Center
Institute of Transportation Studies, UC Berkeley
2614 Dwight Way, Mail code #7374
Berkeley, CA 94720-7374
Email: lizhibin@berkeley.edu

David RAGLAND
Director, Safe Transportation Research & Education Center
Institute of Transportation Studies, UC Berkeley
2614 Dwight Way, Mail code #7374
Berkeley, CA 94720-7374
Tel:510-642-0655
Email: davidr@berkeley.edu

Xiaohong CHEN
Professor, School of Transportation Engineering, Tongji University
Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University
4800 Cao'an Road, Shanghai, 201804, P. R. of China
Tel) +86-21-65989270
Fax) +86-21-65982897
Email: chenxh@tongji.edu.cn

Word Count: 6383(Text) + 500(2 Figures) + 500(2 Tables) = 7383 words
Submission date: Aug. 1, 2012

Abstract

It is widely known that the road network layout can impact the non-motorized users' traffic safety by changing the non-motorized traffic volume and road users' behavior. Different road network patterns lead to different traffic safety levels for non-auto users and a single pattern can even have both the safe and unsafe features at the same time. By knowing what features can lead to safer traffic environment, existing road networks can be improved and new network patterns can be produced by combining all safe features from different patterns. Therefore, the associations between road network structure and pedestrian-bicyclist crashes are analyzed in this paper to determine how the structural features of a road network affect non-motorist safety. Three structural measures including average geodesic distance, network betweenness centrality, and overall clustering coefficient are calculated based on the road networks of 321 census tracts in Alameda County, California. Then the three measures together with other factors like traffic behavior, land use, transportation facility, and demographic features are employed separately in a spatial statistical model called geographically weighted regression. Conclusions are: if a network is more highly centered on major roads, there will be fewer non-motorist crashes; the network which has more average number of intersections between each pair of roads tends to have fewer accidents for pedestrians and bicyclists; and, the more a network is clustered into several sub-core networks, the lower the non-motorist crash count will be.

Keywords

Road network structure, pedestrians, bicyclists, accidents

1 INTRODUCTION

In what pattern roads are connected to each other——the "structure"——determines how direct a route is for vehicles to follow, and how many or what kinds of turns vehicles make in a route. In addition, road network patterns appear to be the dominate influence on travel distance and mode choice which result in different level of attractiveness to pedestrians and bicyclists (1). Thus road network patterns can impact travel behavior and non-motorist volume, which lead to a change in frequency and severity of non-motorist collisions.

1.1 Background

Traffic safety of different road network patterns has always been the major concern of transportation planners and traffic engineers. In 1950s, the accident rates were first compared between grid pattern and curvilinear pattern. On one hand, it showed that the grid pattern had substantially higher accident rate than limited-access pattern (2). Although this study may have "several limitations including control of variables", a series of recent studies using statistical models still imply that discontinuous networks like "loops and lollipops" perform safer than grid iron pattern (3, 4). Two newer studies show the cul-de-sac networks appear to be much safer than the uniform grid networks, by nearly three to one (5), and the grid pattern is found to be the least safe by a significant margin with respect to all other street patterns (6). On the other hand, recent studies have found higher traffic fatality rates in outlying suburban areas than in central cities and inner suburbs with smaller blocks and more-connected street patterns (7, 8, 9). These studies prove that road network pattern can significantly impact traffic safety, although different conclusions have been drawn. Thus, how road network characteristics can effect traffic safety should be investigated rather than just considering the whole pattern.

Structural features, rather than metrical features, focusing on the connection relationships and principles of roads make a pattern different than others. Thus, many studies have been conducted to investigate road network structures. Qualitative studies try to graphically describe road network structures into different categories, such as "grid iron", "fragmented parallel", "warped parallel", "loops and lollipops", and "lollipops on a stick", which is the widely accepted classification method in road network pattern analysis (10, 11, 12). Other studies try to describe network structure quantitatively based on the node-link relationship of a network. Urban planners have developed the conception of connectivity to describe how well a road network links locations, using indices like connected node ratio, Garma index, link-node ratio, etc (13). Network analysts apply topological measures to quantify road networks. Centrality analysis originates in structural sociology, and has been recently introduced to study road systems (14). Limited research on road network centrality show that centrality indices nicely capture the "skeleton" of the urban structure (15), and these indices can allow extended visualization and characterization of the road network structure (16). Other topological measures like network clustering coefficient and geodesic distance are all useful to describe the structure of a network (17). To build the relationship between road network structural features and traffic safety, quantitative measures should be applied in the analysis. The connectivity features have been investigated in Zhang et al. (18), thus in this paper, the topological measurements are utilized.

Recognized as the most vulnerable road users, pedestrians and bicyclists are frequently the focus of traffic safety research. Multitudes of factors have been included in analyses, including vehicle characteristics, roadway design characteristics, road user behaviors, and environmental conditions (19, 20, 21, 22). As road network structures can directly determine the distance and directness of non-motorist's daily trip (1), it is necessary to consider the road network structure as a predominant factor of non-motorist traffic safety. Recent work has begun to investigate the effect of street pattern and compactness on the severity of crashes involving vulnerable road users (23).It shows "loops and lollipops" increases the probability of an injury for pedestrians and bicyclists but reduces the probability of fatality and property-damage-only in an event of a crash. Rather than knowing which pattern is safe for pedestrian, making clear what structural features make a pattern having fewer crashes or more fatal accidents will be more useful.

1.2 Study Objectives

Based on the review of past research, road network patterns can significantly impact traffic safety, but the safety effects of different pattern types are still under debate. Furthermore, being a predominant factor to affect pedestrian-bicyclist volume and driving behavior, the network structural features of a pattern could accordingly lead to different levels of safety for non-motorists travel. Studies about road network structure have offered quantitative measures to make the investigation of association between road network structure and pedestrian-bicyclist accident possible.

Considering the stated issues, the aims of this paper are to examine the relationships between structural characteristics of road networks and pedestrian-bicyclist accidents. Toward this goal, this paper analyzes data from Alameda County, California, at the census tract level which is a proper unit for non-motorized travel study (13). Three measurements, including average geodesic distance, network betweenness centrality, and overall clustering coefficient, from network typology are applied to describe the structural of road network patterns. A spatial statistical model called geographically weighted regression (GWR) is utilized to evaluate the relationship between each structural measure and non-motorist accidents. Within these models, other factors are all included, like travel behavior, transportation facilities, demographic features, and land use.

2 DATA

2.1 Data Source

There are six categories of data employed in this paper, extracted from the road networks, crash records, census statistics, and traffic forecasting models available for Alameda County, California. All the data are calculated and aggregated by census tracts because: first, the median size of census tracts in Alameda County resembles a proper area for walking (trips are typically under one mile) and cycling (trips are typically under 5 miles) (13), with the third quartile value of tract size as 1.01 square miles and 95% tracts are under 5.86 square miles; traffic analysis zones and other spatial units are either too small or too large; second, there are 321 census tracts, which is a proper sample size for GWR models (24). All the data in the study are collected for the same period of time when possible, except the census data which is from the year 2000. However, it is the closest time to satisfy other data and the population structure is always assumed to have not changed much.

Crash Data

Road accidents involving pedestrians and bicyclists in Alameda County, CA from 2004 to 2006 are analyzed in this research. The crash data is from “Transportation Injury Mapping System (TIMS)” which was established by researchers at the Safe Transportation Research and Education Center (SafeTREC) at the University of California, Berkeley. TIMS provides data based on crash records from the “Statewide Integrated Traffic Records System” (SWITRS), and offers mapping analysis tools and information for traffic safety related research, policy and planning. All the crashes are already geocoded on the road network.

Road Network Structure Data

The road network structure characteristics are all calculated based on the road network data using the methods described in the following section. Each census tract has its own structural measures for the road network, and then all these measures together with other regional characteristics will be incorporated into a statistic model.

Travel Behavior Data

Travel behavior data are collected to describe traffic condition from two angles: the first is to use vehicle miles traveled (VMT) to reflect the traffic intensity of the road network in each census tract. This data is obtained from “Bay Area Simplified Simulation of Travel, Energy and Greenhouse Gases” model for 2006, already aggregated by traffic analysis zones and census tracts. Then, the numbers of workers using private vehicles or public transportation or non-motorized means are applied to show the travel mode choices of each area. These data are obtained from U.S. Census 2000 data from the U.S. Census Bureau website.

Land Use Data

Shown to have significant impacts on non-motorized travel (25), the number of commercial units and house units in each census tract are selected to control for the land use impact. The commercial data is from “the Alameda County pedestrian intersection crossing volume model” (ACPICVM) established by SafeTREC. The house unit data is directly from the Census 2000 data. The “year structure built” data are also included in the analysis to reflect the age of an area, calculated as how many house units are built before 1950, because it is indicated in a research that an area built before 1950’s has a different safety performance than areas built in more recent times (26).

Demographic Data

The populations aged from 0 to 15, 16 to 64, and 65 and older are employed to show the population structure; median household income and employment rate are chosen to reflect the economic condition. All these data are from the Census 2000 data.

Transportation Facility Data

The numbers of bus lines in each census tract are aggregated to reflect the transit accessibility, and this is also from the “ACPICVM” mentioned above. Additionally, 3-way, 4-way, more-than-4-way intersection numbers, and connectivity measure such as street densities are calculated based on the road network which is derived from ESRI

"StreetMap North America". Because this paper focuses on the pedestrian-bicyclist crashes, all the primary highway road lines with limited access are excluded.

2.2 Calculation of Road Network Structural Measures

Simplify the Road Network into Topological Network

Structural characteristics focus on the relationship between roads——connected or not. Thus, road networks should be simplified to topological networks which only include nodes and links. There are two ways to obtain a topological network: the primal approach and the dual approach (15, 27). The former is based on a quite simple, intuitive representation of networks which turns intersections into nodes and roads into edges; the latter is opposite by turning roads into nodes and intersections into edges. All the topological measures calculated based on topological network quantify the features of nodes. Since this paper focus on the features of roads rather than intersections, the dual approach is proper to obtain the topological networks. Details about dual approach can be found in the study of Zhang et al (16).

Average Geodesic Distance

To know how far each road is from other roads, one particular definition is the geodesic distance. This quantity is the number of links in the shortest possible route from one node to another. In a topological network, the geodesic distance between two nodes is the count of the number of links in the shortest path between them. When the road network is simplified using dual approach, the geodesic distance between two nodes will be the distance between two roads.

To compare different networks from the perspective of size and efficiency, average geodesic distance is better than the individual one. The average geodesic distance is the ratio of the total geodesic distance of each node pair to the total number of node pairs, as shown in equation 1.

$$GD_{avg} = \sum_{j}^{n} \sum_{k}^{n} \frac{g_{jk}}{n(n-1)/2}, j \neq k \quad (1)$$

Where GD_{avg} is the average geodesic distance of a network, n is the number of nodes in the whole network, g_{jk} is the number of geodesics linking point j and k, $n(n-1)/2$ is the total number of node pairs. A small average geodesic distance suggests a road network in which one road is likely to reach every road through much fewer intersections in between.

Network Betweenness Centrality

Centrality measurements including degree, betweenness and closeness could quantify how 'central' or important each node or link is inside a network (28), so that these measures are appropriate to describe the difference between pattern types. Among all the centrality measurements, the network betweenness centrality index is the best to distinguish different types of road network structure (16), thus this paper use it to describe the structure of a network from the centrality perspective.

The betweenness of a point is "based on the frequency with which a point falls between pairs of other points on the shortest paths connecting them" (28). The higher the betweenness is the more possible a point can fall on the connection path between other points to control their communication. The degree of a point is defined by

$$C_i^B = \sum_{j}^{n} \sum_{k}^{n} \frac{g_{jk(i)}}{g_{jk}}, i \neq j \neq k \quad (2)$$

Where C_i^B is the betweenness centrality of the point i; $g_{jk(i)}$ is the number of geodesics linking point j and k that contain point i on them; g_{jk} is the number of geodesics linking point j and k; $\frac{g_{jk(i)}}{g_{jk}}$ is the probability that point i falls on a randomly selected geodesic linking point j and k; $\sum_{j}^{n} \sum_{k}^{n} \frac{g_{jk(i)}}{g_{jk}}$, the overall betweenness centrality of the point I, is the sum of point i's partial betweenness values for all other pairs of points excluding point i.

Since this paper plans to analyze the centrality property of a whole network, network centralities are applied. The network centralities are based on the point centralities, so there still will be three kinds of them: the network degree centrality, the network betweenness centrality, and the network closeness centrality, all defined by

$$C^B = \frac{\sum_{i=1}^{n} [C_{i*}^B - C_i^B]}{\max \sum_{i=1}^{n} [C_{i*}^B - C_i^B]} \quad (3)$$

Where C^B is the network betweenness centrality; C_i^B is the betweenness centrality of point i defined above; C_{i*}^B is the largest value of C_i^B any point could get in the network; $\sum_{i=1}^{n}[C_{i*}^B - C_i^B]$ is an observed sum of differences to every point's maximum value, and $\max \sum_{i=1}^{n}[C_{i*}^B - C_i^B]$ consequently define the possible maximum sum of these differences. Thus, C^B is defined as "the average difference between the relative centrality of the most central point and that of all other points" (28).

A higher value of network betweenness centrality presents a network which has more roads become the only connection of other roads. This means that there are some roads more central and important than others. For example, according to a recent research about road network centrality (16), grid iron pattern tends to have lower value of network betweenness centrality because every road in this network are equally important to have the same chance to connect to others; cul-de-sac pattern tends to have the higher value of this index, showing that there are some roads overwhelmingly central, connecting almost all of the other roads like a stem of a tree.

Overall Clustering Coefficient

The tendency a large network can be centered toward local sub-networks can be shown as the thought of "clustering" (17). The local clustering coefficient of a node quantifies how close its neighbor nodes are to be in a clique (sub-network). The neighbor nodes of a single node are the nodes which can directly linked to the specific node. The local clustering coefficient is given as follows.

$$CC_i = \frac{\sum_j^m \sum_k^m l_{jk}}{m(m-1)/2}, i \neq j \neq k \qquad (4)$$

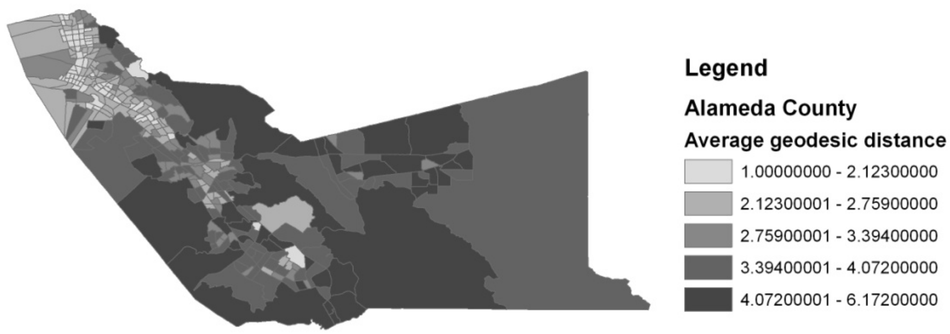
Where CC_i is the local clustering coefficient of node i; m is the number of nodes which are defined as the neighbor nodes of i; j and k are the nodes in the m neighbor nodes of node i; l_{jk} is the link between the m neighbor nodes, if the link exists, $l_{jk} = 1$, otherwise 0 (29).

The overall clustering coefficient for the whole network is given as the average of the local clustering coefficients of all the nodes (29):

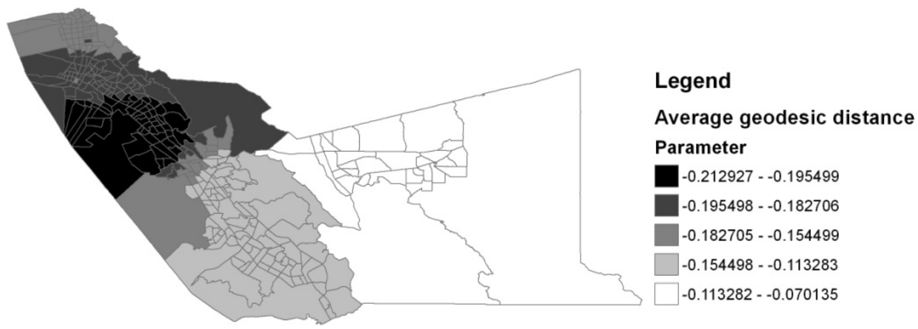
$$CC = \frac{1}{n} \sum_1^n CC_i \qquad (5)$$

If a network has higher overall clustering coefficient, there tend to be more clusters consisting the whole network. For the road network measured in this paper, higher overall clustering coefficient means some roads are highly clustered to several sub-networks in the whole network. In a sub-network, roads are connected directly and efficiently with each other, but not be so with road outside the sub-network.

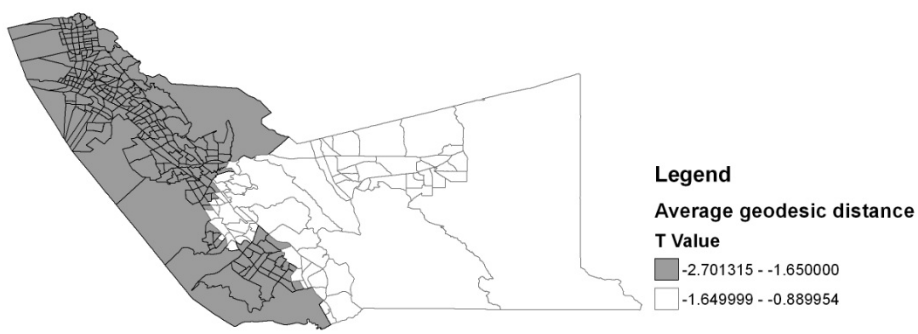
Average geodesic distance, network betweenness centrality, and overall clustering coefficient are structural measurements which can describe a network from the perspective of efficiency, centrality and clustering. For each census tract, there are three measures for its road network. They are calculated by a social network analysis tool called "UCINET" (30). Values are various across the 321 census tracts in Alameda County as shown in figure 1(a), (d), (g) and summarized in table 1.



(a)



(b)

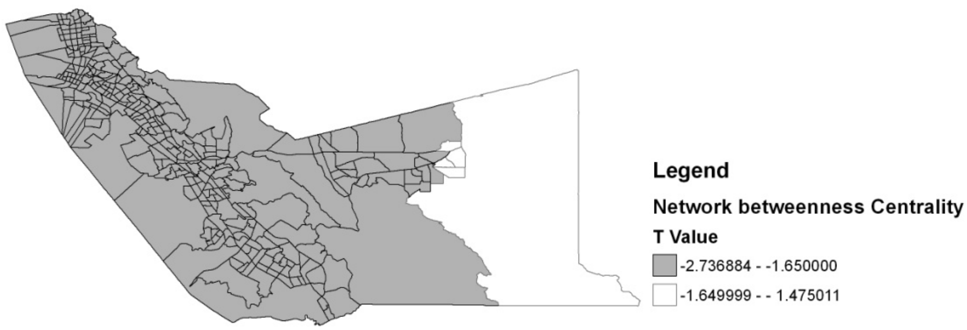


(c)

FIGURE 1 Distribution of the structural measure value and the estimation results: (a) distribution of average geodesic distance, (b) parameter distribution of average geodesic distance, and (c) t value distribution of average geodesic distance.

(d)

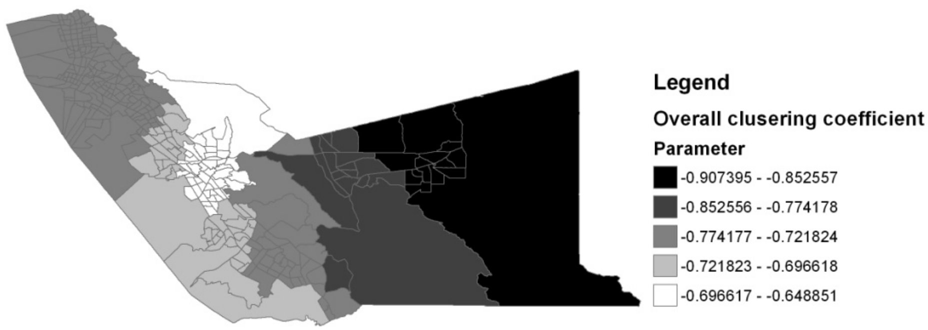
(e)



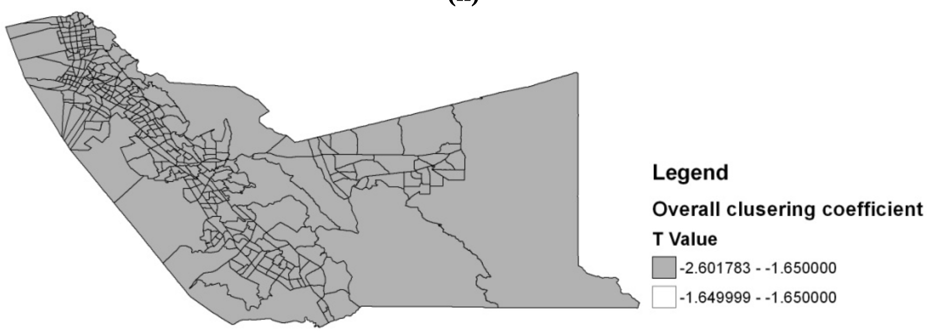
(f)

FIGURE 1 (continued) Distribution of the structural measure value and the estimation results: (d) distribution of network betweenness centrality, (e) parameter distribution of network betweenness centrality, and (f) t value distribution of network betweenness centrality.

(g)



(h)



(i)

FIGURE 1 (continued) Distribution of the structural measure value and the estimation results: (g) distribution of overall clustering coefficient, (h) parameter distribution of overall clustering coefficient, and (i) t value distribution of overall clustering coefficient.

3 STATISTICAL MODELS

Along with descriptive statistics, different statistical models are employed to quantify the relationship between road network features and crash occurrence. The crash data is a type of count data exhibiting over-dispersion so negative binomial regression models have been widely employed to evaluate the association between urban forms and crashes (31). This common technique assumes a spatial stationarity in the relationship between collision count and contributing factors. Under this assumption, fixed coefficients are estimated to represent all the different analysis units for the entire study area, assuming the relationship between dependent variable and independent variables does not vary across the geographic area. However, this stationary relationship may be broken when applying to crash analysis. Safety performance is likely influenced by many factors which are spatially defined and related between continuous areas such as census tracts, traffic analysis zones, or census blocks. These factors could be land use, demographic features, and road networks, which could be strong predictors at some locations but weak at others (31). For example, when the relationship between crashes and intersection numbers is estimated for each census block in a region, the estimation result could be different across census blocks with different income level. For low income level locations, more intersections could expose cars in more conflicts, thus there could be more crashes. However in other locations with higher income level, the number of intersections may not have a significant impact on crashes because residents with high income can afford expensive vehicles which have better safety protection to potentially offset the increase of crashes. As a result, ignoring the spatial non-stationarity between crashes and spatial related factors could lead to the inaccuracy of model findings.

3.1 Introduction to Geographically Weighted Regression

To address the non-stationarity problem mentioned above, geographically weighted regression (GWR) has been developed to allow relationships between dependent and independent variables to vary across locations (24). Consider a regular regression model written as:

$$y_i = \beta_0 + \sum_k \beta_k x_{ik} + \epsilon_i \quad (6)$$

Where y_i is the dependent variable observed in location i; β_0 is the interception; k is the total number of independent variables; β_k is the parameter of the kth independent variable; x_{ik} is the kth independent variable observed in location i; ϵ_i is the error term for the estimation in location i. β_k is estimated globally and do not change with locations so that this model is called "global" model.

GWR allows local rather than global parameters β to be estimated by extending this traditional regression framework as:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \epsilon_i \quad (7)$$

Where (u_i, v_i) denotes the coordinates of the ith location point (census tract centroid in this study) in the study area; $\beta_k(u_i, v_i)$ is a realization of the continuous function $\beta_k(u, v)$ at location I, so GWR models can be called "local" models compared to the traditional ones. In this way the GWR recognizes the existence of spatial variations in relationships and calibrates the equation in a reasonable way—weighted regression. For the purpose of this paper is not to introduce the calibration of GWR, detailed information about calibration could be found in relative research (24), and the calculation in this paper will be finished using a software called GWR 3.0 (32).

3.2 Model Specification

Model Form

The basic GWR assume a normally distributed error structure in the calibration of the regression model. This assumption is not upheld when calibrating models for count data so a Poisson distribution is thus more appropriate. Although a negative binomial distribution is better than the Poisson distribution because of the over-dispersion of crash data, the use of Poisson regression does not produce inaccurate estimates (31). Furthermore, considering the availability of Poisson regression for GWR 3.0 software utilized in this study, the model of this study is developed using the Poisson distribution form as:

$$Ln(y_i) = \beta_0(u_i, v_i) + \beta_1(u_i, v_i) Ln(Exposure) + \beta_2(u_i, v_i) x_{i2} + \cdots + \beta_k(u_i, v_i) x_{ik} \quad (8)$$

Where "Exposure" is the exposure variable in Poisson regression model; others are the same as mentioned above.

Variables

The dependent variable is the average crashes involving pedestrian and bicyclist per year. It is calculated as the mean of the crashes from 2004 to 2006 to minimize the data fluctuation through years. The independent variables are classified into five categories: structural measures, land use, travel behavior, transportation facilities, and demographic features, as shown in table 1. This paper chooses population density instead of VMT as the exposure variable because previous research at the TAZ level indicates that VMT does not perform well as exposure when the study unit is a continuous area rather than individual facility. Also, as one of the widely used exposures, population density also can reflect strong positive relationship with traffic crashes especially for regional study (33).

1 **TABLE 1 List of Variables and the Model Structure**

Category	Variable	Symbol	Avg	Min	Max	S.D.	Selection of Variables in 15 models #														
							1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Dependent variable: crashes	Number of crashes involving pedestrians and bicyclists for each census tracts per year. (based on crashes from 2004-2006)	y_i	4.01	0	24	3.49	√														
Exposure variable	Population density(persons/mile^2)	PD	10077.20	7154.03	26.70	38852.20	√														
Structural measures	Average geodesic distance	GD_{avg}	2.99	1	6.17	0.91	√ (select one measure for each model)														
	Network betweenness centrality	C^B	34.61	2.56	79.74	14.44															
	Overall clustering coefficient	CC	0.27	0	0.68	0.15															
Land use	Number of Commercial properties in the census tract	ComCnt	22.11	0.00	142.00	23.24		√				√	√	√				√	√		√
	Number of housing units in the census tract	HUTot	1682.81	10.00	4969.00	751.53															
	Rate of house units built before 1950	E50	0.33	0.00	1.00	0.27															
Travel behavior variables	Vehicle miles traveled	TotVMT	56465.35	288.36	228925.30	36628.58			√			√			√	√		√	√	√	√
	Number of workers 16 years and over who go to work using private vehicle	WTPRV	1695.17	6.00	5557.00	963.02															
	Number of workers 16 years and over who go to work using public transportation	WTPUB	224.84	0.00	1090.00	163.26															
	Number of workers 16 years and over who go to work using biking or walking	WTBW	94.40	0.00	1389.00	143.39															
Transportation facility variables	Number of bus lines in the census tract	BusCnt	62.79	0.00	316.00	54.60				√			√		√		√	√		√	√
	Number of 3 way intersections in the census tract	3WayIntCnt	63.34	4.00	837.00	64.75															
	Number of 4 way intersections in the census tract	4WayIntCnt	28.34	4.00	131.00	16.05															
	Number of more-than-4-way intersections in the census tract	MWayIntCnt	1.17	0.00	9.00	1.41															
	Street density (miles/ mile^2)	StDen	20.85	1.04	38.08	8.27															
Demographic variables	Population age 0 to 5	Pop15	991.11	0.00	2878.00	579.65					√			√		√	√		√	√	√
	Population age 16 to 64	Pop16_64	3046.74	23.00	8885.00	1394.88															
	Population age 65 and older	Pop65	459.79	1.00	1478.00	247.90															
	Employment rate	EmR	0.47	0.10	0.68	0.10															
	Average household income in 1999	HHInc	59060.48	2499.00	167106.00	27426.64															

2

Model Structures

Considering the co-linearity between different measures, all 3 structural measures will be employed separately in series of models. And because there are so many independent variables that a forward procedure is used in this paper to test which variables should be included in a model to make the best estimation (31). In this procedure, a simple model with only a structural measure, an exposure variable, and an intercept term is used as a starting point. Then, other control variables will be added to the model one category by one category. This procedure produces 15 models. In each model, there will be one structural measure, together with other control variables, as shown in table 1. Also, prior to incorporating variables into the same model, a correlation test has been conducted to examine whether variables are highly correlated with each other. If two variables are substantially correlated, they will not be inserted into the same model.

4 RESULTS

4.1 Parameter Estimation for Structural Measures

GWR calibrates local models for each location, so that the results of the GWR models are a set of local parameters for each independent variable. Therefore, each variable will have 321 estimations for parameter, t value, and standard error, varying across 321 census tracts. Focusing on the impacts of structural, the parameters for each structural measure in different models are summarized in table 2. Since each structural measure can have 321 parameters estimated in each model, the parameters are presented in the order of the minimum, the lower quartiles, the median quartiles, the upper quartiles, and the maximum values from top to bottom in each cell in table 2.

TABLE 2 Parameter Estimations for 3 Structural Measures in Different Models

Structural measure	Parameters estimated for each structural measure in model #1 to #8							
	#1	#2	#3	#4	#5	#6	#7	#8
Average geodesic distance	-0.781415 -0.460909 -0.164200 -0.021454 0.317146	-0.301509 -0.289797 -0.256382 -0.099239 -0.063505	-0.577583 -0.363115 -0.235386 0.018223 0.222258	-0.431982 -0.202221 -0.097281 -0.018606 0.089000	-0.485216 -0.280714 -0.212170 -0.051829 0.153911	-0.126515 -0.088150 -0.070981 -0.066453 -0.046162	**-0.212927** **-0.191759** **-0.179498** **-0.132799** **-0.070135**	-0.132216 -0.127987 -0.118755 -0.074298 -0.030167
Network betweenness centrality	-0.031982 -0.018180 -0.013580 -0.007995 0.008544	-0.011594 -0.011433 -0.010802 -0.006916 -0.005318	-0.029711 -0.016538 -0.009144 -0.004879 0.006415	-0.015995 -0.009054 -0.006596 -0.005623 -0.001076	-0.016335 -0.009102 -0.007071 -0.004218 0.002931	**-0.009127** **-0.007665** **-0.006983** **-0.006561** **-0.006360**	-0.008367 -0.007957 -0.007781 -0.006227 -0.005382	-0.006700 -0.006486 -0.005905 -0.002484 -0.001475
Overall clustering coefficient	-2.502191 -1.549860 -1.172694 -0.585080 1.112635	-1.153597 -1.099589 -1.084213 -0.922471 -0.860129	-2.553119 -1.630452 -1.261362 -1.002860 0.352393	-1.723497 -1.271504 -0.876937 -0.578707 0.145134	-1.390687 -0.981670 -0.734158 -0.431692 0.705764	-0.930012 -0.870591 -0.828093 -0.802433 -0.780891	-1.027092 -0.976087 -0.948353 -0.891902 -0.856620	-0.756083 -0.726075 -0.639400 -0.355112 -0.264480
Structural measure	Parameters estimated for each structural measure in model #9 to #15							
	#9	#10	#11	#12	#13	#14	#15	-
Average geodesic distance	-0.165117 -0.087571 -0.064341 -0.025300 0.039935	-0.298942 -0.194911 -0.147574 -0.017314 0.018388	-0.150478 -0.084158 -0.057092 0.014291 0.138831	-0.157208 -0.116508 -0.088476 -0.072909 -0.012383	-0.061353 -0.044666 -0.026599 -0.013220 0.000669	-0.018718 0.001111 0.016766 0.033354 0.091793	-0.078672 -0.049056 -0.020564 -0.000479 0.033253	-
Network betweenness centrality	-0.012290 -0.007322 -0.004062 -0.002705 -0.001839	-0.007442 -0.007153 -0.005557 -0.005303 -0.004237	-0.005121 -0.003567 -0.003440 -0.003076 -0.002346	-0.006393 -0.005663 -0.004676 -0.003778 -0.003167	-0.002966 -0.002655 -0.002577 -0.001705 -0.000932	-0.004843 -0.002205 0.000954 0.002022 0.002624	-0.002650 -0.002022 -0.001782 -0.001388 -0.001065	-
Overall clustering coefficient	-1.519654 -1.168403 -1.102005 -0.730529 -0.223090	-1.088696 -1.054784 -0.419880 0.032745 0.736365	-0.583514 -0.560169 -0.483137 -0.249945 -0.065760	**-0.907395** **-0.734046** **-0.731246** **-0.717829** **-0.648851**	-0.440218 -0.420971 -0.361835 -0.189965 -0.087669	-0.460213 -0.406280 -0.269540 -0.036737 0.236939	-0.459516 -0.424627 -0.341853 -0.083263 -0.020493	-

4.2 Significance Level

The t statistic value for each parameter also varies across census tracts as shown in figure 1, so there is no single t value to represent the significant level of the estimation as it is expected in the regular global regression. Thus, this paper develops the "significance rate" value—calculated as the rate of census tracts whose parameter estimations for structural have t value bigger than 1.65 or smaller than -1.65 (±1.65 is the critical value for two tailed t-test at the confidence level of 90% for degree of freedom of 320)—to show how many census tracts have statistically significant relationship between structural measures and non-motorist accidents at 90% confidence level. The significance rate for parameter estimation of each structural measure is shown in figure 2 along the horizontal axis titled as "level of significance". This paper assume that an estimation is fair if the significance rate is higher than 0.6. Therefore, the points on the right side of the black vertical line represent models which have better ability to estimate parameters at 90% confidence level.

4.3 Better-fit models

Regarding how effectively the model can describe the relationship, this paper introduces the Corrected Akaike Information Criterion (AICc) (24): a lower value of AICc indicates a better fit model. The AICc values for all the models are displayed in figure 2, from the minimum value of 234.89 to the maximum value of 536.97. This paper assumes that the AICc value lower than 350 indicates a better-fit model. As can be seen in figure 2, the points under the black horizon line represent models which have better fitness.

In order to further study the associations between each structural measure and non-motorist accidents, models which can effectively describe the relationship with more significant estimations should be selected for analysis. Therefore, both the significance level and goodness of fit are considered together. First, the models with AICc value lower than 350 are focused on. Then, among these lower AICc models, those for each structural measure with the highest significance rates are picked out for further analysis: model No.7 with the structural measure of average geodesic distance; model No.6 with the structural measure of network betweenness centrality; and model No.12 with the structural measure of overall clustering coefficient, as shown in bolded cells in table 2 and solid black points in figure 2.

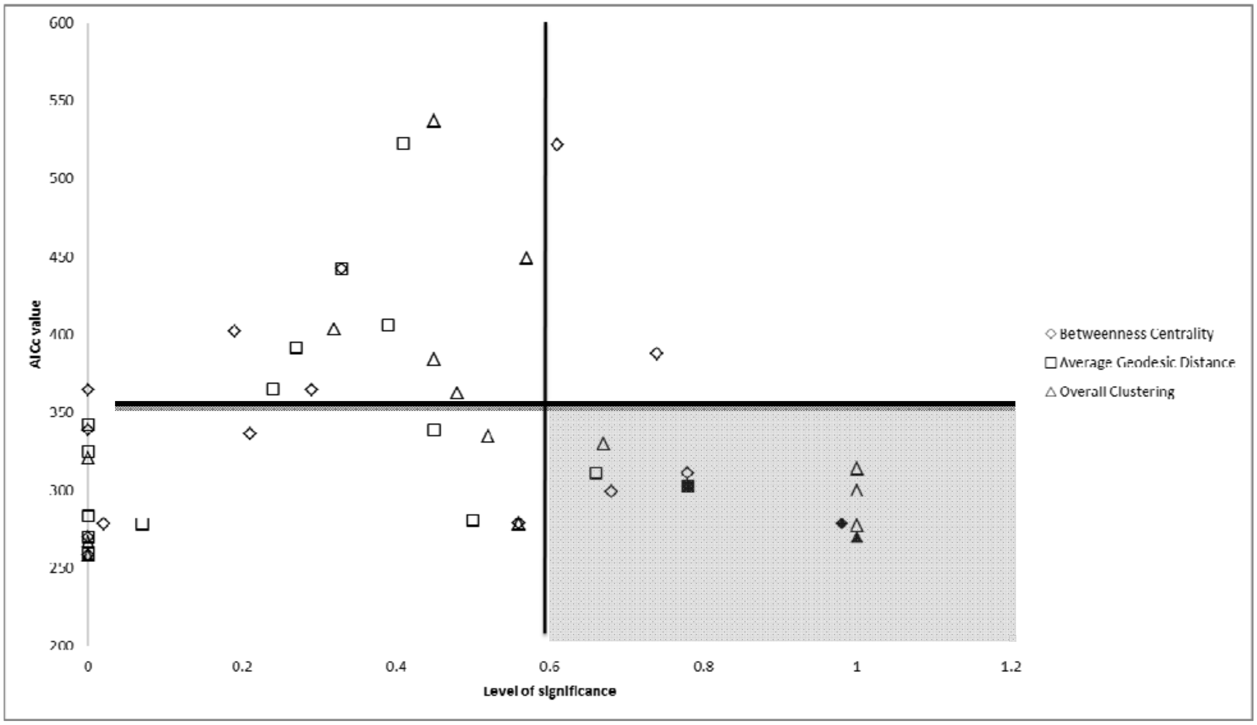


Figure 2 The significance level of parameter estimation for structural measures and goodness of fit for different models.

5 CONCLUSIONS

According to the significance rate and model fitness level, 3 models are chosen to examine the associations between road network structure and pedestrian-bicyclist accidents.

5.1 Positive Or Negative?—The Direction of Impact

The GWR results show the parameter of a certain variable is not constant for all census tracts. Thus the parameter for structural measure will vary across different census tracts, as shown in figure 1.

Figure 1 (b) shows the parameter distribution of average geodesic distance among all the census tracts, with negative values in all tracts. At the same time, the conclusion should be drawn together with the comparison to the t value distribution of this measure. As mentioned above, the t values are different for a variable in different locations. Shown in figure 1 (c), the t values for parameter estimations of average geodesic distance vary across the 321 census tracts. The areas with t value bigger than -1.65 provide insignificant parameters estimation, shown as the white areas in the east part of Alameda County in figure 1 (c). It is obvious that the tracts which have significant parameter estimations are all with negative parameters. Thus, a significant negative relationship between average geodesic distance and pedestrian-bicyclist crashes can only be concluded in the west part of Alameda County, which means larger average geodesic distance is related to a decrease of non-motorist crashes.

In figure 1 (e), the parameter estimations for network betweenness centrality are all negative values that range from -0.009 to -0.006. And these estimations are significant at 90% level in the east and middle part of

Alameda County. So in these areas, it can be concluded that network betweenness centrality has a negative relationship with pedestrian-bicyclist crashes: the more centered a network, the safer for pedestrians and bicyclists.

In figure 1 (i), the parameter estimations for network betweenness centrality are all significant at 90% level for 321 census tracts. And the parameters are all negative from -0.9 to -0.6 shown in figure 1 (h). So it can be concluded that overall clustering coefficient has a negative relationship with pedestrian-bicyclist crashes: the more clusters a network has, the fewer accidents for pedestrians and bicyclists.

5.2 Weak or Strong?—The Degree of Impact

The parameters estimated by GWR not only vary in signs, but also change in quantities, which indicate the degree of the influence of a variable can be different across tracts. The effects could be strong in some census tracts but be weak in others. Regardless of whether the sign is negative or positive, the absolute values of a parameter show how strong the structural measure is related to pedestrian-bicyclist safety. As shown in figure 1 (b), the parameters of average geodesic distance in the east part of Alameda County can be very close to 0, which means that the influence of this measurement on safety can be very weak for these tracts. When one unit of increase happens to the average geodesic distance, no noticeable change could be expected for accident number. However, every conclusion should be drawn at certain confidence level. Comparing the t value distribution in figure 1 (c), it is easy to find that the small absolute values in figure 1 (b) are all the tracts with t values bigger than -1.65 which means insignificant estimation at 90% confidence level. In other words, in highly significant tracts the absolute values of parameters are relatively large.

Shown in figure 1 (e), the parameters for network betweenness centrality are with stronger impacts (higher absolute values) in the middle of Alameda County and weaker impacts (lower absolute values) in the north west. Compare the figure 1 (f), the lower absolute values in east are ignored because the estimations here are insignificant at 90% confidence level. From figure 1 (h) and figure 1 (i), overall clustering coefficient can strongly influence the non-motorist accidents in the east, becoming weaker from the out range to the core in the county.

6 DISCUSSION

Based on the stated conclusions, it is clear that higher structural measures means safer environment for pedestrians and bicyclists. Higher average geodesic distance would indicate a network with roads that are further apart from each other. In other words, one has to pass more intersections when traveling between origins and destinations located on any two roads if a network has high average geodesic distance, which also implies an inefficient network to avoid through traffic for vehicles. Less vehicle volume could be the reason for less pedestrian-bicyclist crashes. Larger network betweenness centrality means roads in a network highly possibly centered onto some specific main roads, so that these main roads become center roads as trunks with others as branches. To understand this kind of network, it is likely that the branch roads may only serve local traffic and the trunk roads will serve through traffic. If the network betweenness centrality is high, there will be more branch roads centered onto fewer trunk roads. Then this network can also discourage through traffic, which leads to fewer conflict possibility between cars and non-motorists. A higher overall clustering coefficient indicates that a network is highly clustered, or that the network contains many clusters in which roads are well-connected between themselves, but badly connected to other clusters. Thus this kind of network tends to have sub-groups of roads as neighbors in a huge community. Then, the majority of these roads will serve as local roads in a neighbor more than as collector roads. Again, this kind of network will have less through traffic for cars, which offers a safer environment for pedestrians and bicyclists.

Comparing the road network patterns and their structural measures, it can be inferred that grid iron patterns tend to have lower average geodesic distance, network betweenness centrality, and overall clustering coefficients, and that the grid iron pattern may lead to more non-motorist crashes. On the other hand, a study about non-motorist traffic safety impact of road network connectivity shows higher connectivity, such as denser street, higher block density, and shorter mean block length, can lead to reduction of accidents for pedestrians and bicyclists (18). If investigating the road network patterns and their connectivity measures, it can be inferred that grid iron pattern is more likely to have higher connectivity. Thus, one could conclude that grid iron pattern is safer for pedestrians and bicyclists. At first glance, the conclusions are contrary. But that is only because we draw the conclusion too abruptly. First, a network which has lower structural measures is not necessarily a grid iron pattern. Or more precisely speaking, a network with lower structural measures does not necessarily have higher connectivity measures. For example, there could be a road network pattern has higher overall clustering coefficient and higher street density at the same time to make it safer for pedestrians and bicyclists. Second, even if in most cases it is correct to infer that an efficient, less centered, and clustered network is grid iron network, which means the safety impacts of grid iron pattern are contrary in the two studies, it still does not necessarily mean any of the study is wrong. It only means one certain pattern of road network can have good and bad effects on non-motorist safety at the same time. And this is the reason why we need to investigate what features of a network that can lead to good impacts and combine all the good features from different patterns to make a new design such as the fused grid pattern (6). Third, safety impact for a road network can be different depending on what crash data is applied in analysis. For example, if fatal crash counts is used, it may conclude that more four-way intersections can be safer; but if it is the total crashes considered, the four-way intersections may turn out to be an unsafe factor (34). Further studies should consider the relationship between road network measures and different patterns, and their safety impacts on pedestrians and bicyclists should be investigated with different crash severities.

REFERENCE

1. Vicky Feng Wei, Gord Lovegrove. Sustainable road safety: A new neighborhood road pattern that saves VRU lives. *Accident Analysis and Prevention*, Vol. 44, 2012, pp. 140-148.
2. Michael Southworth, and Eran Ben-Joseph. *Streets and the shaping of towns and cities*. Mcgraw-Hill, 1996.
3. Shakil Mohammad Rifaat, Richard Tay, and Alex de Barros. Effect of Street Pattern on Road Safety: Are Policy Recommendations Sensitive to Aggregations of Crashes by Severity? In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2147, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 58-65.
4. Eran Ben-Joseph, Livability and Safety of Suburban Street Patterns: A Comparative Study (Berkeley, CA: Institute of Urban and Regional Development, University of California, Working Paper 641, 1995).
5. Gordon R. Lovegrove and Tarek Sayed. Using Macrolevel Collision Prediction Models in Road SafetyPlanning Applications.In *Transportation Research Record: Journal of the Transportation Research Board*,No. 1950, Transportation Research Board of the National Academies, Washington,D.C., 2006, pp. 73–82.
6. Sun, J. & Lovegrove, G. Research Study on Evaluating the Level of Safety of the Fused Grid Road Pattern, External Research Project for CMHC, 2009, Ottawa, Ontario.
7. Ewing R, Schieber RA, and Zegeer CV. Urban Sprawl as A Risk Factor In Motor Vehicle Occupant And Pedestrian Fatalities. *American Journal of Public Health*, Vol. 93, No. 9, 2005, pp. 1541-1545.
8. Pawlovich, M., Li, W., Carriquiry, Alicia et al. Iowa's Experience with Road Diet Measures: Use of Bayesian Approach to Assess Impacts on Crash Frequencies and Crash Rates. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1953, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 163-171.
9. Wesley E. Marshall, and Norman W. Garrick. Street Network Types and Road Safety: A Study of 24 California Cities. *Urban Design International*, Vol. 15, No. 3, 2010, pp. 133-147.
10. Fanis Grammenos, and Julie Tasker-Brown. Residential street pattern design. *Socio-economic Series 75*, 2002. http://www.cmhc-schl.gc.ca/publications/en/rh-pr/tech/socio75.html , Accessed Jun 1st, 2010.
11. Shakil Mohammad Rifaat, and Richard Tay. Effects of Street Patterns on Injury Risks in Two-Vehicle Crashes. Presented at 88th Annual Meeting of the Transportation Research Board,Washington, D.C., 2009.
12. Michael Southworth, and Eran Ben-Joseph. Reconsidering The Cul-de-sac. *Access*, No. 24, University of California, 2004.
13. Jennifer Dill. Measuring Network Connectivity for Bicycling and Walking. Presented at 83rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2004.
14. Paolo Crucitti, Vito Latora, and Sergio Porta. Centrality Measures in Spatial Networks of Urban Streets. *Physical Review E*, Vol. 73, No. 036125, 2006.
15. Sergio Porta, Paolo Crucitti, and Vito Latora. The Network Analysis of Urban Streets: A Primal Approach. *Environment and Planning B: Planning and Design*, Vol. 33, No. 5, 2006, pp. 705 – 725.
16. Yuanyuan Zhang, Xuesong Wang, Peng Zeng et al. Centrality Characteristics of Traffic Analysis Zone Road Network Patterns. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2256, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 15-24.
17. Hanneman, Robert A. and Mark Riddle. 2005. Introduction to social network methods. Riverside, CA: University of California, Riverside (published in digital form athttp://faculty.ucr.edu/~hanneman/)
18. Yuanyuan Zhang, John Bigham, et al. Associations between road network connectivity and pedestrian-bicyclist accident. Presented at 90th Annual Meeting of the Transportation Research Board, Washington, D.C., 2012.
19. Mohamed Abdel-Atya**,** Sai Srinivas Chundia and Chris Lee. Geo-Spatial and Log-Linear Analysis of Pedestrian and Bicyclist Crashes Involving School-Aged Children. *Journal of Safety Research*, Vol. 38, No. 5, 2007, pp. 571-579.
20. Eluru N, Bhat CR, and Hensher DA. A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes. *Accident Analysis and Prevention*, Vol. 40, No. 3, 2008, pp. 1033-1054.

21. Kim JK, Kim S, Ulfarsson GF et al. Bicyclist Injury Severities in Bicycle-Motor Vehicle Accidents. *Accident Analysis and Prevention*, Vol. 39, No. 2, 2007, pp. 238-251.
22. Macpherson AK, To TM, Parkin PC et al. Urban/Rural Variation in Children's Bicycle-Related Injuries. *Accident Analysis and Prevention*, Vol. 36, No. 4, 2004, pp. 649-654.
23. Rifaat SM, Tay R, and de Barros A. Effect of Street Pattern on the Severity of Crashes Involving Vulnerable Road Users. *Accident Analysis and Prevention*, Vol. 43, No. 1, 2011, pp. 276-283.
24. A. Stewart Fotheringham, Chris Brunsdon, Martin Charlton. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons Ltd, England, 2002.
25. Robert J. Schneider, Lindsay S. Arnold, and David R. Ragland. Validation Testing and Refinement of the Alameda County Pedestrian Intersection Crossing Volume Model White Paper. Association of Collegiate Schools of Planning Conference, 2009.
26. Amrosell. Improve Safety By Increasing Street Connectivity. http://tcstreetsforpeople.org/node/652, July 2011.
27. Sergio Porta, Paolo Crucitti, and Vito Latora. The Network Analysis of Urban Streets: A Dual Approach. *Physica A*, Vol. 369, 2006, pp. 853-866.
28. Linton C. Freeman. Centrality in Social Networks Conceptual Classification. *Social Networks*, Vol. 1, 1978. pp. 215-239.
29. D. J. Watts and Steven Strogatz. "Collective dynamics of 'small-world' networks". *Nature*, Vol. 393 (6684), 1998. Pp. 440–442.
30. Borgatti, S.P., Everett, M.G. and Freeman, L.C. Ucinet for Windows: Software for Social Network Analysis. 2002, Harvard, MA: Analytic Technologies.
31. Alireza Hadayeghi, Amer S. Shalaby, Bhagwant N. Persaud. Development of Planning Level Transportation Safety Tools Using Geographically Weighted Poisson Regression. *Accident Analysis and Prevention*, Vol. 42, 2010, pp. 676-688.
32. GWR 3.0 Software Package. Martin Charlton, Stewart Fotheringham, Chris Brunsdon. Spatial Analysis Research Group. Department of Geography, University of Newcastle upon Tyne, Newcastle upon Tyne, ENGLAND, 2009.
33. Felipe Ladron de Guevara, Simon P. Washington, and Jutaek Oh. Forecasting Crashes at the Planning Level: Simultaneous Negative Binomial Crash Model Applied in Tucson, Arizona. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1897, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 191-199.
34. Eric Dumbaugh and Robert Rae. Safe Urban Form: Revisiting the Relationship Between Community Design and Traffic Safety. *Journal of the American Planning Association*, Vol. 75, No. 3, Summer 2009