THE CRASH INTENSITY EVALUATION USING GENERAL CENTRALITY CRITERIONS AND A GEOGRAPHICALLY WEIGHTED REGRESSION

M. Ghadiriyan Arani [a] , P. Pahlavani [b,*], M. Effati [c] , F. Noori Alamooti [d]

[a] M.Sc. student, GIS Division, School of Surveying and Geospatial Eng, College of Eng, University of Tehran - m.ghadiriyan@ut.ac.ir
[b] Assistant Professor at School of Surveying and Geospatial Eng., College of Eng., University of Tehran - pahlavani@ut.ac.ir
[c] Assistant Professor at Department of Civil Eng, University of Guilan - meysameffati@guilan.ac.ir
[d] MSc Graduate, Department of Civil Eng, Shahid Rajaee Teacher Training University - s_n9194@yahoo.com

KEY WORDS: Crash intensity, General centrality criterions, Geographically weighted regression, Atlanta highway

ABSTRACT:

Today, one of the social problems influencing on the lives of many people is the road traffic crashes especially the highway ones. In this regard, this paper focuses on highway of capital and the most populous city in the U.S. state of Georgia and the ninth largest metropolitan area in the United States namely Atlanta. Geographically weighted regression and general centrality criteria are the aspects of traffic used for this article. In the first step, in order to estimate of crash intensity, it is needed to extract the dual graph from the status of streets and highways to use general centrality criteria. With the help of the graph produced, the criteria are: Degree, Pageranks, Random walk, Eccentricity, Closeness, Betweenness, Clustering coefficient, Eigenvector, and Straightness. The intensity of crash point is counted for every highway by dividing the number of crashes in that highway to the total number of crashes. Intensity of crash point is calculated for each highway. Then, criteria and crash point were normalized and the correlation between them was calculated to determine the criteria that are not dependent on each other. The proposed hybrid approach is a good way to regression issues because these effective measures result to a more desirable output. R^2 values for geographically weighted regression using the Gaussian kernel was 0.539 and also 0.684 was obtained using a triple-core cube. The results showed that the triple-core cube kernel is better for modeling the crash intensity.

1. INTRODUCTION

The driving crashes are one of social dilemmas causing the death of a large number of people and imposing the heavy costs to each society in the world especially for developing countries. To identify the crash intensity points within the city together with marginal information in order to improve the safety level of transportation networks for allocating the resources is a necessary task. The crash intensity points are a part of a route that they have potential for hazardous and crash due to some factors and conditions. Mainly, the prioritization methods of crash intensity points are based on single-criterion approaches. These criteria can be mentioned such as the number of crashes, crash severity, similarities in crashes, and financial damages. The different factors play role in emerging a crash and determining the crash intensity points that in such conditions, it should be studied the spatial structure for urban routes. In fact, the spatial structure is a way which is effective on passages placement. In this study, the objective is to investigate the effect of network structures on within-city crashes. As it is expected, the problem of effect of network structures on crashes follows two parts of spatial structure and within-city crashes. The spatial structures in transportation networks consist of the arrangement and layout of all parts of a network. Daily increase vehicles and travel within the urban has been too increased in urban areas. To reduce crashes, crash analyses are needed that several studies have been conducted in recent years. From the perspective of spatial networks (Jiang et al., 2008, 2011; Jiang, 2008) define the concepts of configuration and spatial structure and the principles and concepts of spatial configuration and simulation and modelling of spatial structure in networks. Newman et al. (2006) compared the spatial structure transport networks in different cities. Shu (2009) and Iida et al. (2005) have tried to use the

concepts related to spatial structure networks in an application such as crisis management and criminology. About urban crashes Levine et al. (2003) investigated the role of spatial analysis as an all-round way in road crashes and proper management of roads. In Hong Kong, the cartographic analysis and point analysis were used to show the pattern of road crashes (Chin Lai et al., 2008). In Maku County in the state of Michigan in America, an artificial neural network model has been used to predict the number of crashes have been occurred at intersections (Akin et al., 2010). The criteria that are included in this research for analysis of crashes are general centrality criteria that have traffic concepts. These criteria include Degree, Pageranks, Random walk, Eccentricity, Closeness, Betweenness, Clustering coefficient, Eigenvector, and Straightness.

Degree criterion on graph refers to the number of input node to major node (Barrat et al., 2004), since in the real world more connected to a street intersection, it is more likely to crash. Closeness criterion is the reverse total distance of the geodesic major node from other existing node in the network. The geodesic distance of two nodes in the network means the shortest distance between two nodes (Freeman et al., 1997). Betweenness criterion was proposed by Freeman et al. (1997). This criterion shows that how much a node is involved in shortest routes between different parts of the network. Clustering coefficient criterion indicates the willingness of nodes on network for producing a cluster (Opsahl et al., 2009). Pageranks criterion are the key technology in the Google search engine, that is calculated based on a return relationship, so that a node that has a higher rank causes the nodes connected to it also have a higher rank (Brin et al., 1998). Eigenvector criterion is the Degree criterion improved, it considers quality of connections in addition to their amounts in evaluating a node (Spielman, 2007). Randomwalk criterion is the study of the properties of routes made up by

random and sequential steps of a motion in the study area (Blanchard et al., 2009). Straightness criterion specifies that how much diversion exists between the connecting path and Euclidean distance between two nodes (Latora et al., 2001). Eccentricity criterion represents the maximum distance between a node to the other nodes (Mislove, 2009). Before running the algorithm, correlations between criteria mentioned above should be investigated. For this purpose, the covariance and correlation between two sets X and Y is used as follows:

$$cov(X.Y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n} \qquad (1)$$

$$r = \frac{cov(X.Y)}{\sigma_x \sigma_y} \qquad (2)$$

where $\bar{x}.\bar{y}$ is the average of data sets, n is the number of every set with deviation of $\sigma_x.\sigma_y$, and r shows the correlation between X, Y. High correlation-near to 1 shows that two sets are interdependent and both of them cannot be used in calculations. As shown in Figure 1, the correlation between criteria is in a range [-0.6, 0.6] that shows theses criteria do not have a significant correlation with each other. Hence, all of them are used in the algorithm.

ID	CRITERIA	ID	CRITERIA	ID	CRITERIA
1	Degree	4	Betweeness	7	Random walk
2	Eigenvector	5	Straightness	8	Clustering coefficient
3	Closness	6	Pageranks	9	Eccentricity

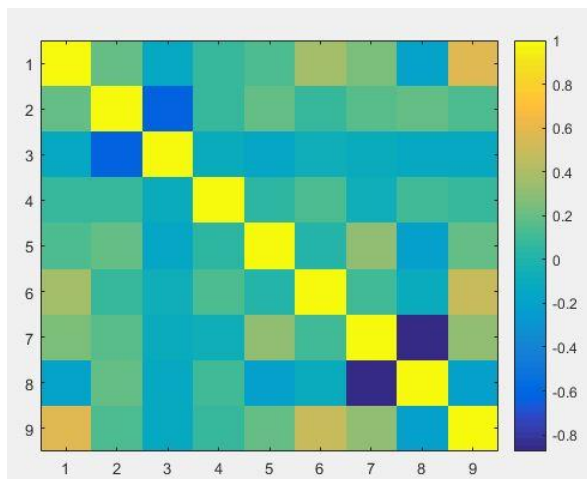Table 1. The criteria used in this paper

Figure 1. Correlation matrix for general centrality criterions

2. METHODOLOGY

2.1 Generating the training data

For generating the training data, any crash point was allocated to one highway. For each highway, there is a need to specify criteria. In this regard, the dual graph of highways was formed at first. In dual graph, each line represents a point and each point defines a line such as Figure 2. The specified criteria that should be achieved by the dual graph are listed in Table 2.

Line	Degree	Pageranks	Random walk	Eccentricity	Eccentricity	Betweeness	Clustering coefficient	Eigenvector	Straightness	Crash
1	4	2	0.3	0.6	2	4	0.2	0.6	0	0.127

Table 2. A sample of the training data criteria
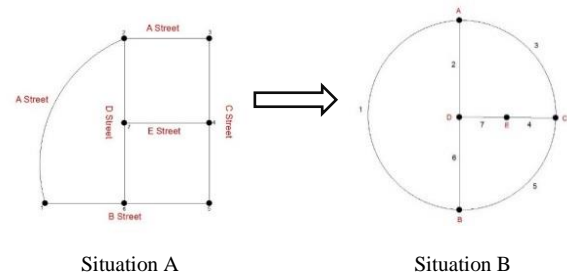

Situation A   Situation B

Figure 2. Converting the major graph-situation A to dual graph-situation B

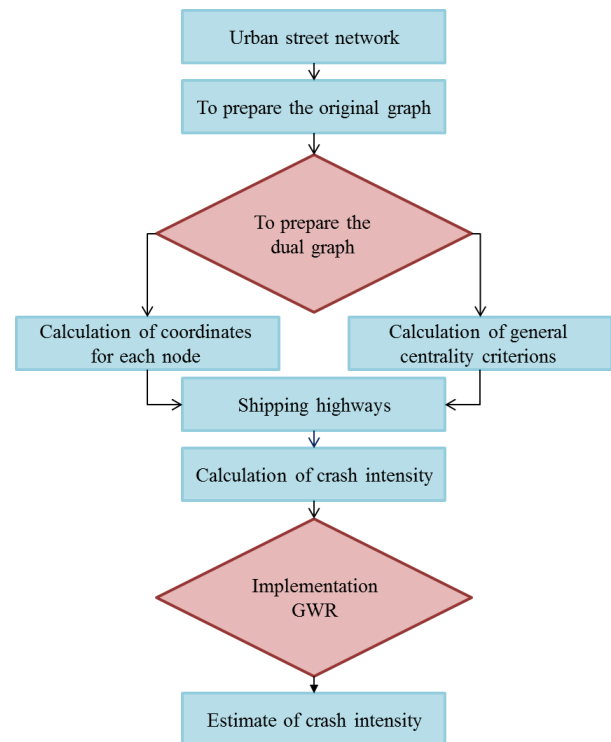Figure 3 shows the process of preparing data to predict crash intensity and extraction of crash intensity.

Figure 3. Solving process

2.2 Geographically weighted regression

The basic regression model has a series of pre-assumptions that one of them is data independency, but the local data has certain characteristics that are difficult to work with them. Two samples of these characteristics include (a) local autocorrelation Butler's law defining the inverse relationship or distances (Tobler, 1970)

, and (b) local non-stationary that represents a change in local autocorrelation in space and environmental heterogeneous.

The local autocorrelation may exist between variables or other model characteristics. This means that neighbour variables may be the same value or if we draw the remains of model on a map, the magnitude and location of the remaining symbols are the same neighbours. Existing of local autocorrelation between the remains of model leads to an inefficient estimation. The standard errors of the parameters will be too large. There is also a local structure of the data indicating the dependent variable in a local unit is under effect of values of the independent variables in adjacent local units, and this issue leads to inefficiency in addition to the bias generation. This means that estimates will be too small or large.

In 1988, Anselin presented two models to deal with the issues. A local model that is suitable for the local autocorrelation between remains and other, the delay model which is suitable for the local autocorrelation between data (Anselin, 1988). When the maximum likelihood estimation is used, the parameter estimation can be done for both models without bias.

Another phenomenon is the heterogeneity of the environment that we face in local modelling. It is usually assumed that regression models show the relationship between the variables identical in under study area. This assumption is known as of homogeneity environment. But the different problems such as the various data generation methods violate this assumption and in this case we encounter with problem of heterogeneous environments.

The first models developed to deal with the problem of heterogeneous environments is the expansion method (Casetti, 1972). In This model, the parameters are a function of location and can be written as a polynomial in terms of spatial coordinates. Then, by using the method of least squares can be calculated the unknowns in the model. An important part of this method is to select the right degree of polynomials in order to model that requires an understanding of the variables nature and it will affect the results, significantly.

Geographically weighted regression (GWR) allows different relationships to exist at different points in the study area and improves the modelling performance by reducing spatial autocorrelations. In addition, these relationships also greatly depend on scale, which is inherent in natural and man-made processes and patterns (Lu et al., 2001). Therefore, local rather than global parameters can be estimated, and spatial non-stationarity can be detected at multi-scales by changing bandwidth of GWR. Geographically weighted regression were used to investigate the relationships between landscape fragmentation and related factors. Since OLS is well known, we will give only a brief introduction for the theoretical background of the GWR model in the next paragraphs. Moreover, steps for data pre-processing and stationary index calculation are also described in this section.

The conventional global regression can be expressed as:

$$\hat{y} = \beta_0 + \sum_{j=1}^{p} \beta_i X_j + \varepsilon \tag{3}$$

where $\hat{y}$ is the estimated value of the dependent variable at location j , β_0 represents the intercept, β_i expresses the slope coefficient for independent variable X_j is the value of the variable X_j at location i, and ε denotes the random error term for location i. In this equation, the estimates of the model parameters are assumed to be spatially stationary.

The GWR model extends conventional global regression by generating a local regression equation for each observation, and the above model can be rewritten as:

$$\hat{y} = \beta_0(u_i, v_i) + \sum_{j=1}^{p} \beta_i(u_i, v_i) X_j + \varepsilon \tag{4}$$

where (u_i, v_i) denotes the coordinate location of the ith point, $\beta_0(u_i, v_i)$ is the intercept for location i , $\beta_i(u_i, v_i)$ represents the local parameter estimate for independent variable X_j at location i.

Parameter estimates in GWR are obtained by weighting all observations around a specific point i based on their spatial proximity to it. The observations closer to point i have higher impact on the local parameter estimates for the location, and are weighted more than data far away. The parameters are estimated from:

$$\hat{\beta}(u,v) = (X^T W(u,v) X)^{-1} X^T W(u,v) y \tag{5}$$

where $\hat{\beta}(u,v)$ represents the unbiased estimate of β , $W(u,v)$ is the weighting matrix which acts to ensure that observations near to the specific point have bigger weight value.

The weighting function, called the kernel function, can be stated using the exponential distance decay form:

$$W_{ij} = \exp\left(\frac{d_{ij}^2}{b^2}\right) \tag{6}$$

where W_{ij} represents the weight of observation j for location i , d_{ij} expresses the Euclidean distance between points i and j, and b is the kernel bandwidth. If observation j coincides with i, the weight value is one. If the distance is greater than the kernel bandwidth, the weight will be set to zero.

3. RESULT

In this section, the data, dual graph and criteria, as well as an example, in addition to the structuring of geographically weighted regression are discussed.

3.1 Data

Atlanta is the capital and the most populous city in the U.S. state of Georgia, with an estimated population of 463,878 in 2015. Atlanta is the cultural and economic center of the Atlanta metropolitan area, home to 5,522,942 people and the ninth largest metropolitan area in the United States. Atlanta is the county seat of Fulton County, and a small portion of the city extends eastward into DeKalb County. In Atlanta, due to the large number of highways, there are many potential risks. Every year, a significant number of crashes occur on these roads. This town has 24 highways that all of them have been used in this study to calculate and predict crash intensity for every highway. Figures 8 and 9 show the position of the town and Figure 10 shows the current situation highways in town.
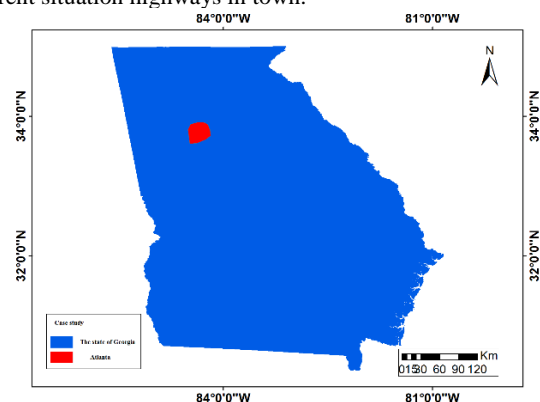

Figure 4. Position of the study area
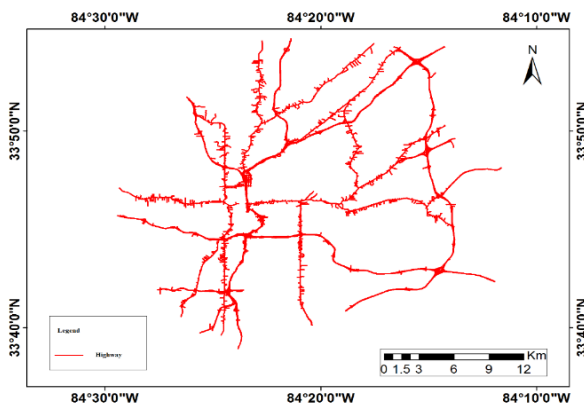
Figure 5. Position of the study area

Figure 6. Crash intensity with triple-cube kernel

Figure 6. Current situation of highways
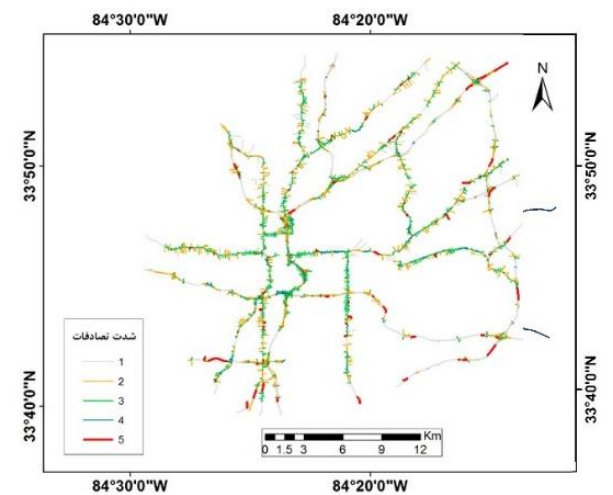
Figure 7. Crash intensity with Gaussian kernel

3.2 Dual graph and criteria

The first step is to build the dual graph. Then, the criteria for each highway that would be converted to a node in the dual graph are calculated. 9 criteria as well as the geographical coordinates, besides the crash intensity for every node should be calculated. These 9 criteria and geographical coordinates are used as the inputs for GWR and the output of GWR is the intensity of crash for every node.

3.3 Structuring the GWR

The criteria and crash intensity were normalized and formed as inputs to the designed GWR. Determining the geographic weights is very important. For this reason many cores have been proposed. Two well-known of these cores have been proven high performance including the Gaussian kernel and the triple-cube kernel. The designed GWR will output the computed intensity of the crash point by numbers in the range [-1, 1] that -1 shows less intensity for a point and +1 shows the highest intensity for a point. Afterwards, they were re-scaled and classified to the five degrees of intensity crash that 5 shows a higher intensity. The classified map of the study is presented in the Figures 6 and 7.
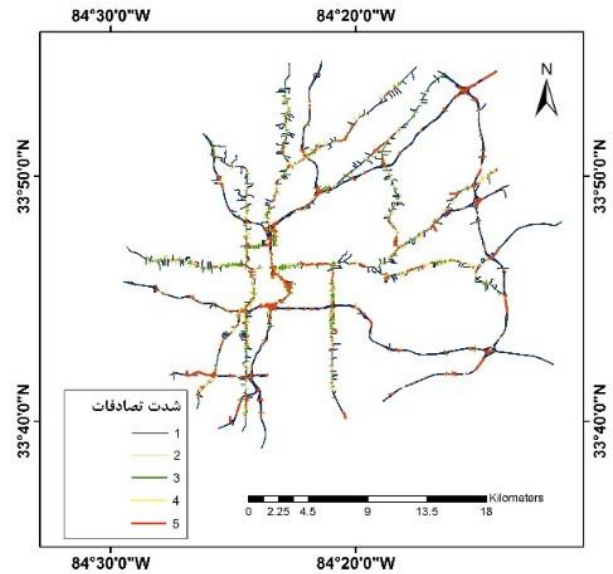
Table 3 presents the values of R^2 and RMSE used in two cores.

Approach	R^2	RMSE
Triple-cube kernel	0.684	0.0865
Gaussian kernel	0.539	0.0312

Table 3. Accuracy criteria based on two cores

As results shown in Figure 6, only 10% of highways in this region have a high degree of crash intensity. As shown in Figure 6, most of the lines are 2 and 1 degree. Also, the higher degree of crash intensity resulted by some criteria types such as Degree, Clustering coefficient and Eigenvector. The results of implementing the GWR in Atlanta showed that the current situation of highways is in accordance with living standards in urban environments.

4. CONCLUSION

In the urban land development and control planning, highway should be considered as the basis for the policies and planning

projects that are attempting to mitigate urban land use problems and lead to a sustainable development in highway. In this regard, assessing the current situation and addressing the problems to be considered in highway crash intensity evaluations is a major task in urban land development and control planning. This research aimed to develop a new model for modeling the degree of crash intensity using the general centrality criteria of highways. In this regard, GWR was designed to model and predict the intensity of crash for a highway using the general centrality criteria. This regression classifies the impacts of each highway to reflect a crash intensity. Assessment of crash intensities in Atlanta showed the need for long-time plans to make an acceptable balance among related concerns of different highways in this town. As a result, the degree of crash intensity map achieved through the proposed GWR can define the current problematic areas of an urban environment that enables urban planners to regulate and control the changes and improvements in land via creating policies to support sustainable development.

REFERENCES

Akin D., B. A. (2010). "A neural network (NN) model to predict ntersection crashes based upon driver." Scientific Research and Essays 19: 10.

Anselin L., "A test for spatial autocorrelation in seemingly unrelated regressions," Economics Letters, vol. 28, pp. 335-341, 1988.

Barrat, A., et al. (2004). "The architecture of complex weighted networks." Proceedings of the National Academy of Sciences of the United States of America 101(11): 3747-3752.

Blanchard, P. and D. Volchenkov (2009). Mathematical analysis of urban spatial networks, Springer.

Brin, S. and L. Page (1998). "The anatomy of a large-scale hypertextual Web search engine." Computer networks and ISDN systems 30(1): 107-117.

Casetti E., "Generating models by the expansion method: applications to geographical research," Geographical analysis, vol. 4, pp. 81-91, 1972.

Chin Lai P., W. Y. C. (2008). "GIS for road crash analysis in hong kong." JGIS 10: 10.

Freeman, L. C. (1977). "A set of measures of centrality based on betweenness." Sociometry: 35-41.

Iida, B. H. a. S. (2005). "Network and psychological effects in urban movement." Lecture Notes in Computer Science 3693.

Jiang B., C. L. (2008). "Street-based topological representations and analyses for predicting traffic flow in GIS." International Journal of Geographic Information Science.

Jiang B., T. J. (2011). "Agent-based simulation of human movement shaped by the underlying street structure." International Journal of Geographical Information Science 25: 25.

Jiang, B. (2008). "Flow dimension and capacity for structuring urban street networks." Statistical Mechanics and its Applications 387: 12.

Latora, V. and M. Marchiori (2001). "Efficient behavior of small-world networks." Physical review letters 87(19): 198701.

Levine, N., Karl E. Kim, and Lawrence H. Nitz (2003). "Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns." Crash Analysis & Prevention 11: 11.

Lu, Y. H., & Fu, B. J. (2001). Ecological scale and scaling. Acta Ecologica Sinica, 21(12), 2096-2105, (in Chinese).

Mislove, E.A. (2009) Online Social Networks: Measurement, Analysis, and Applications to Distributed Information System (Huston, Texas).

Newman, M. G. a. M. (2006). "The spatial structure of networks." The European Physical Journal 49: 5.

Opsahl, T. and P. Panzarasa (2009). "Clustering in weighted networks." Social networks 31(2): 155-163.

Shu, C. (2009). Spatial Configuration of Residential Area and Vulnerability of Burglary, in 7th International Space Syntax Symposium, Stockholm: KTH

Spielman, D. A. (2007). Spectral Graph Theory and its Applications. FOCS.

Tobler W. R., "A computer movie simulating urban growth in the Detroit region," Economic geography, vol. 46, pp. 234-240, 1970.