Analysis of Metro Manila Road Network Robustness through Reachability Matrix

Leandro Isla[a], Kardi Teknomo[a]

[a]*Ateneo de Manila University, Loyola Heights, Quezon City,, Philippines*

Abstract

This study analyzes the robustness of a road network under natural and man-made disasters through an index introduced by the authors based on the reachability matrix. As our main case study, we analyze the road network of Metro Manila by simulating destruction of random locations on the network. The road network used in this study was created using data from the Open Street Map mirror site, Metro Extracts. Results of this study show that the Metro Manila road network is fairly robust up to 1 km radius of destruction, where a maximum of 4.86% paths disrupted is incurred. However, the damage sharply increases for larger radii of destruction. Five percent or more loss in the network's connectivity is noted in 13.8% of the nodes for the 2 km radius and in 77.3% of the nodes for the 4 km radius.

Keywords: network robustness; reachability matrix; network measures; disaster management; disaster mitigation

1. Introduction

The disruption of a city's road network has a great range of effects, from minor inconveniences to significant loss of access to basic commodities and services. This study seeks to analyze the robustness of the Metro Manila road network by observing the effects of the destruction of random areas on the paths of the network. The relationships among the damage incurred in the network's connectivity, the number of roads destroyed, and the length of roads destroyed are also determined.

Metro Manila is a bustling metropolis in the Philippines that spreads across 638.6 square kilometers with a population of 11.8 million out of the nationwide total of 92.3 million as of the 2010 census [1]. Lying in the middle of two major fault-lines and a typhoon path, Metro Manila is the world's second riskiest 'city', behind Tokyo-Yokohama in Japan, according to a 2013 poll by Swiss Re, an international reinsurance company [2]. As the national center of government, commerce, and education in the Philippines, any disaster that renders a significant portion of the city unreachable would certainly have a negative impact on the country.

A measure called the Reachability Index, denoted by RI, is introduced here. The RI varies between 0 and 1, with 1 meaning the network is a strongly connected graph where every node has a path to every other node and 0 meaning that there are no edges on the network.

2. Related Literature

2.1. Network Measures

Transportation systems are commonly represented using networks as an analogy for their structure and flows.

Several measures are used to define the structural attributes, specifically the connectivity, of a road network, among them the Beta and Gamma indices. Connectivity here is an attribute of a network that measures the minimum number

of edges needed to reach all nodes from all other nodes [3]. The Beta Index β measures the level of connectivity in a graph and is expressed by the relationship between the number of edges e over the number of nodes v; i.e., $\beta = \frac{e}{v}$. Trees and simple networks have Beta value of less than one. A connected network with one cycle has a value of 1, while more complex networks have a value greater than 1.

The Gamma Index γ is a measure of connectivity that considers the relationship between the number of observed edges and the number of possible edges; i.e., $\gamma = \frac{e}{3(v-2)}$ for planar graphs. The value of Gamma is between 0 and 1, where a value of 1 indicates a completely connected network which is extremely unlikely in reality.

The Beta and Gamma indices remain limited in revealing structural differences between two networks of equal size. More robust measures which take into account the internal complexity of the graph have thus been proposed [4].

2.2 Tarjan's Strongly Connected Components Algorithm

Tarjan in 1972 created an algorithm for detecting the strongly connected components of a network [5]. This algorithm will hereafter be named TA in this paper. TA performs a depth-first search on a given graph, subdividing it into groups where each node is strongly connected to each other node inside the group. A graph is said to be strongly connected if the algorithm returns only one group. The resulting graph is also acyclic. TA has been used in the simplification of water distribution system networks [6, 7].

2.3 Network Robustness

There have been several studies on network robustness. Albert et al.[8] define robustness through the analysis of the diameter of the network and the relative size of fragmenting clusters. Callaway et al. [9] use the size of the giant component. Both these studies simulated the removal of random nodes from the graph and observed the effects on specific attributes of the graph.

Latora and Marchiori [10] use the concept of local efficiency to quantify a network's resistance to failure on a small scale, that is, how well the information is exchanged by the neighbors of a disabled node. Scott et al. [11] introduced the Network Robustness Index as a means for evaluating the importance of a given highway segment with regard to the increased travel-time cost of rerouting all other traffic should the segment be removed. This concept was used by Sullivan et al. [12] to identify critical links in Chittenden County, Vermont.

Several definitions of robustness are available in the literature. A well-known definition is given by Ziha [13], who defines robustness as the capacity of a system or network to respond to adverse conditions.

The concept of robustness is closely related to the concept of reliability. Bell and Iida [14] define reliability as the degree of stability of the quality of service which a system normally offers. For Imers et al. [15], the distinction between the two concepts lies in the fact that reliability refers to evaluation of the performance of a system, while robustness is a characteristic of the system itself.

Santos et al. [16] addressed robustness through three reliability measures, namely Network Spare Capacity, City Evacuation Capacity, and Network Vulnerability. For Network Vulnerability, they noted that the failure of some links can have serious consequences on the overall performance of an interurban road network. A network that is less vulnerable to the failure of links is a network that is more robust and vice versa [17].

2.4 Reachability Matrix

As cited in [18], Harary et al. [19] denoted the reachability matrix by N_{p-1}, which is obtained by raising the adjacency matrix N_1 to the $(p$-1)th power, where p is the number of nodes . However, since this process becomes unwieldy due to the large number of nodes used in this study, the authors added a new halt condition. The algorithm stops running when the resulting matrix is filled with ones.

3. Methodology

3.1. Map Creation

The graph of Metro Manila used in this study was taken from Open Street Map (OSM). Two important OSM elements are relevant to this study: a *node* which represents a particular point in the world with its own latitude and longitude and a *way* which is an ordered set of nodes that are implicitly connected by a series of line segments. While ways can be used to represent many things, such as buildings or rivers, the representations relevant to this study are the ways that are highways. An OSM file the size of Metro Manila cannot be downloaded directly from the OSM website [20], so it was necessary to download an extract from a mirror site called Metro Extracts [21]. Metro Extracts makes available weekly updated city-sized extracts for major cities all over the world. It offers these extracts in a variety of formats, such as OSM PBF, OSM2PGSQL SHP, and OSM XML, which was used in this study.

As the authors could not extract the road network directly and perform experiments on it, they decided to create their own free program to perform the map creation, data cleaning, and experimentation used in this study. It is available in the references [22].

The OSM file was processed into two files: Nodes and Road Links. The Nodes file contains each declared node in the map used for highways. Each node has a unique id, and a coordinate composed of a longitude and a latitude. The Road Links file contains the edges of the graph, each identified by the id of its origin node and the id of its destination node. Thus, a two-way road will be represented as two separate edges. The file also contains the length of the edge, calculated as the Haversine distance between the origin and the destination, the number of lanes of the road (defaults to 2 if unspecified as per the OSM Wiki [23]), the type of road, and the name of the road, if available.

There are hundreds of types of roadways in OSM, but most, such as footways, are not relevant to this study. The types of highways used were restricted to Detail Level 5, which consists of Motorways, Trunks, Primary, Secondary, and Tertiary Roads. These types of highways were chosen since these are the most important roads in what OSM terms as a "standard road network" [24]. A Motorway is a fast, restricted access road, normally with two or more lanes plus an emergency hard shoulder. A Trunk road or national highway is where smaller "feeder" roads connect and where traffic tends to congregate. For the readers who are familiar with Metro Manila, the South Luzon Expressway (SLEX) and the Epifanio de los Santos Avenue (EDSA) are examples of a motorway and a trunk road, respectively. Usually, Primary roads link larger towns, Secondary roads link smaller towns and villages, and Tertiary roads connect minor streets to more major roads [22]. OSM classifies shorter segments as link variants, e.g., motorway link. For the purposes of this study, link variants are classified by their parent type, e.g. motorway links as motorways.

3.2. Data Cleaning and Preprocessing

One problem with the Metro Extracts map is the presence of way tags that refer to node ids that are not declared in the file. This leads to gaps in roadways which are supposed to be connected and would affect the experiments, so the missing node ids were collated into a file used to query the OSM API. The node information retrieved from the API was then inserted back into the Metro Manila OSM XML file. The fundamental assumption of this study is that barring any disaster, a car can travel from any point in the city to any other point. After the initial extraction of edges, it was observed that the resulting Metro Manila graph was not strongly connected. This happens because 3% of the edges of the network are isolated, one-way roads near the boundary of the map, or through contributor error. To clean the OSM graph data, we use Tarjan's Strongly Connected Components Algorithm (TA), which goes through the graph in a manner similar to a DFS and organizes the nodes into groups where each node inside the group can reach every other node inside the group. These groups are called *strongly connected components*. If these components are treated as nodes and used to create a graph, the resulting graph is guaranteed to be acyclic. The digraph resulting from the extraction process of Metro Manila contains several hundred components, most of which contain only a single node. Highways with a large width, such as Trunks and Motorways, are often represented as two separate OSM Ways, one for each direction. We illustrate the cleaning process of the map into a fully connected component in Fig. 1. The two one-way roads heading towards the edge of the map are made of multiple components.

In Fig. 1, the nine nodes in the Core Component represent the cleaned, fully connected network. There are two one-way roads, one going toward the core and the other going away from the core. Both roads were cut off by the edge of the map. When TA is performed on this graph example, it will detect seven distinct components. Performing TA on the Metro Manila road network produces the Core Component, which is fully connected.
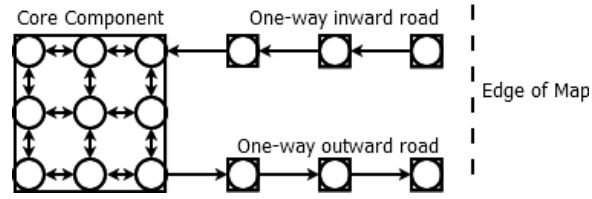
Fig. 1. Example of roads that are removed

We analyze the effect this node and road removal has on the entire road network. Tables 1 and 2 show that the effect is less than 3%. Thus it is reasonable to just take the Core Component and disregard the others. About 97% of the nodes and road network information are concentrated in the Core Component, containing almost the entirety of Metro Manila, especially the downtown areas. This justifies the assumption that the initial Reachability Index of Metro Manila is 1. We note the difference between an Edge and a Road: a Road represents the actual street on the map and may be either one-way or two-way, while an Edge is the traditional Graph Theory Edge where a two-way Road is represented as two separate Edges.

Table 1: Nodes Before and After Cleaning

Nodes Before Cleaning	Nodes After Cleaning	After/Before
56903	55330	**97.24%**

Table 2. Roads Before and After Cleaning

Road Type	MTY	TRK	PRI	SEC	TER	Total
Edges After/Before	93.69%	98.54%	96.70%	98.53%	97.65%	97.62%
Length After/Before	93.60%	97.97%	97.20%	98.41%	96.87%	97.24%
Roads After/Before	93.55%	98.15%	97.47%	98.34%	97.10%	97.47%

3.3 Reachability Index and Condensation

The Reachability Index (RI) is calculated from the Reachability Matrix, which is also called the Path Matrix or the Binarized Path Matrix [25]. An element in an Adjacency Matrix is equal to 1 if there exists an edge from the row node to the column node. For this study, a node is always treated as though it has a self-loop. The Reachability Matrix is a superset of the Adjacency Matrix [25]. Each element is equal to 1 if there exists a path from the row node to the column node, where a path is a series of connected edges from an origin to a destination. The RI is the sum of all elements in the Reachability Matrix divided by the maximum possible sum, which is n^2, where n is the number of nodes. To illustrate this concept, for the fully-connected graph in Fig. 2, RI = 9/9 = 100%.
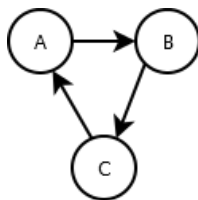

Fig. 2. Simple graph

	A	**B**	**C**
A	1	1	0
B	0	1	1
C	1	0	1

(a)

	A	**B**	**C**
A	1	1	1
B	1	1	1
C	1	1	1

(b)

Fig. 3. (a) Adjacency matrix and (b) Reachability matrix of Fig. 2

Condensation is the process of simplifying a graph by grouping strongly connected nodes together into a "super node" and creating a new graph that uses these "super nodes" as nodes. Nodes are considered to be strongly connected when there exists a path to and from each node in the group. A group of connected nodes is called a *component*. In Fig. 4(a), there are three strongly connected components in the graph: the U group with dashed lines, the V group with gray lines, and the W group with dotted lines. The edges which connect nodes within the same component are called *internal edges* while the ones that connect nodes between components are called *external edges*. In Fig. 4(a), the solid black directed line segments are the external edges. In the process of condensation, all internal edges are removed, and each group is treated as a single node, as illustrated in Fig. 4(b). Condensation is an important part of simplifying the calculation of the RI. To illustrate, the RI of Fig. 4(a) is shown in Fig. 5. Thus, increasing the number of nodes in the map would increase the memory requirements of the Reachability Matrix quadratically. For Fig. 5, RI = 43/64 = 67.1875%.
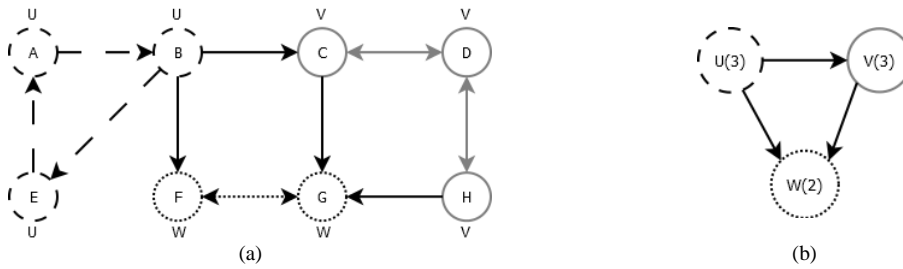

Fig. 4. (a) Sample graph with colored components (b) Condensed Graph

	A	B	C	D	E	F	G	H
A	1	1	1	1	1	1	1	1
B	1	1	1	1	1	1	1	1
C	0	0	1	1	0	1	1	1
D	0	0	1	1	0	1	1	1
E	1	1	1	1	1	1	1	1
F	0	0	0	0	0	1	1	0
G	0	0	0	0	0	1	1	0
H	0	0	1	1	0	1	1	1

Fig. 5. Reachability matrix of Fig. 4(a)

An alternative method of calculating the RI is through the use of the Reachability Matrix of the condensed graph and its Weighted Matrix, which is a matrix where each element is the product of the number of nodes in the corresponding row component and column component of the condensed graph. By performing an element-wise multiplication of the Reachability Matrix in Fig. 6(a) and Weighted Matrix in Fig. 6(b), getting the sum, and dividing the result by n^2, we get RI = (9 + 9 + 6 + 9 + 6 + 4) / 64 = 43 / 64 = 67.1875%, which is the same as the RI of the non-condensed graph. This alternative method is important since there are 55,330 nodes in a Level 5 map of Metro Manila.

	U	V	W
U	1	1	1
V	0	1	1
W	0	0	1

	U(3)	V(3)	W(2)
U(3)	9	9	6
V(3)	9	9	6
W(2)	6	6	4

(a)   (b)

Fig 6. (a) Reachability matrix and (b) Weighted matrix of condensed graph

3.4 Algorithm

The following algorithm was used in calculating the Paths Disrupted, Roads Destroyed, and Length Destroyed in this study. The node id is a randomly selected id in the network. 1000 nodes were selected for each radius. The radii chosen were 0.25 km, 0.5 km, 1 km, 2 km, and 4 km, thus there are a total of 5000 nodes used as centers. The network graph was created from an OSM file of Metro Manila. The Paths Disrupted is 1 – RI. This was done because it is more convenient to say 5% damage rather than dropped to 95% connectivity. The Roads Destroyed is the

number of road segments within the selected radius of the selected node. The Length Destroyed is the sum of the length of all segments destroyed.

Input: node id, radius, network graph
Output: paths disrupted, roads destroyed, length destroyed

1. Select node id in network graph to act as center.
2. Select all road segments within radius from center.
3. Destroy all selected road segments.
4. Calculate RI. Paths Disrupted is 1 – RI. Note the destroyed segments as Roads Destroyed. Length Destroyed is the sum of the lengths of each segment.
5. Return Paths Disrupted, Roads Destroyed, Length Destroyed.

3.5 Experimental Design

Five sets of independent samples, each of size 1000, were chosen randomly, and each set was subjected to a certain radius of destruction. The radii of destruction chosen were 0.25 km, 0.5 km, 1 km, 2 km, and 4 km. The resulting damage to the network's connectivity was measured through the RI. The number and length of roads destroyed were also noted.

4. Results and Analysis

Table 3 shows the mean paths disrupted, roads destroyed, and length destroyed among the five radii. As the radius increases, the mean paths disrupted, mean roads destroyed, and mean length destroyed also increase. The Kruskal-Wallis test showed that there is a significant difference at the 0.01 level in the paths disrupted, roads destroyed, and length destroyed among the five radii, namely, 0.25 km, 0.5 km, 1 km, 2 km, and 4 km. Thus the paths disrupted, roads destroyed, and length destroyed among the five radii were not the same. The Tukey HSD test was run to determine which of the five radii differ significantly in the paths disrupted, roads destroyed, and length destroyed. It showed that at the .05 level, there is no significant difference in the paths disrupted for radii 0.25 km and 0.5 km, while paths disrupted for radii 1 km, 2 km and 4 km are significantly higher than those of 0.25 km and 0.5 km. Also, there are significant differences in the paths disrupted among the 1 km, 2 km, and 4 km radii, wherein larger radii have significantly greater paths disrupted. The same results on significant differences hold for roads destroyed and length destroyed.

Table 3: Test of Difference of Paths Disrupted, Roads Destroyed, and Length Destroyed According to Radius

Radius (km)	Paths Disrupted	Roads Destroyed	Length Destroyed
0.25	0.001842099^a	33.2240^a	1.6432^a
0.50	0.003787935^a	85.9640^a	4.3670^a
1.0	0.010142153^b	250.46^b	12.9151^b
2.0	0.031376533^c	844.44^c	44.4408^c
4.0	0.094153646^d	2679.2^d	144.83^d
Kruskal-Wallis Results	K (Chi-Square) = 4086.155** P value =0.000	K (Chi-Square) = 4177.784** P value =0.000	K (Chi-Square) = 4270.692** P value =0.000

Column means with same letter superscript are not significantly different at the 0.05 level (Tukey HSD)
** Difference is significant at the 0.01 level (2-tailed).

Table 4 shows that there is a significant relationship between paths disrupted and roads destroyed and between paths disrupted and length destroyed, with correlation value (Spearman correlation coefficient) of 0.979 and 0.966, respectively, where each value implies a very strong positive linear correlation. Thus we have sufficient evidence to conclude that as the paths disrupted in the road connectivity increases, the number of roads destroyed and the length of road destroyed also increase. Also, there is a significant relationship between roads destroyed and length destroyed with correlation value (Spearman correlation coefficient) of 0.983 which implies a very strong positive linear correlation. Thus we have sufficient evidence to conclude that as the number of roads destroyed increases, the length of roads destroyed also increases. Fig. 7 below shows the strong linear relationship between paths disrupted and roads destroyed, especially on the bottom portion of the graph.

Table 4: Correlation between Paths disrupted, Links Destroyed, and Length Destroyed

Fig. 8. Histogram of paths disrupted for 0.5 km radius

Fig. 9 shows that the distribution of paths disrupted for 1 km radius is still skewed to the right. Majority of the paths disrupted were below 1.005% which comprises to 59.5%, while only 8.7% of the paths disrupted were above 2.005%. Moreover, the mean paths disrupted is 0.010142153 with a standard deviation of 0.006641467.
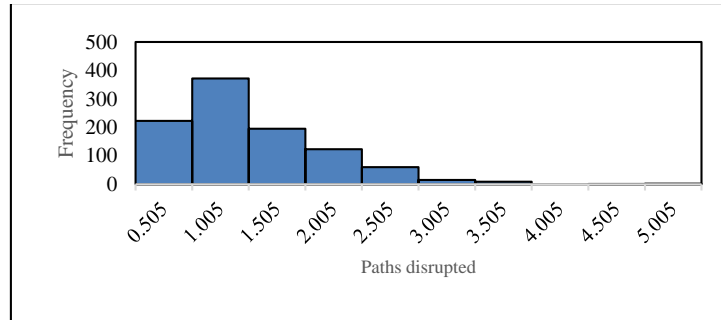

Fig. 9. Histogram of paths disrupted for 1 km radius

For the 2 km radius, 13.8% of the paths disrupted ranged from 5% to 7.91%. Moreover, as shown in Table 7, majority of the paths disrupted were below 4.005% which comprises to 73%, while only 8.9% of the paths disrupted were above 6.005%.

Table 7. Frequency Distribution of Paths disrupted for 2 km radius

Class Intervals (%)	Class Boundaries (%)	Freq.	Cumulative Freq.	Cumulative Relative Freq.
0.01 – 1.00	0.005 – 1.005	80	80	0.08
1.01 – 2.00	1.005 – 2.005	234	314	0.314
2.01 – 3.00	2.005 – 3.005	182	496	0.496
3.01 – 4.00	3.005 – 4.005	234	730	0.73
4.01 – 5.00	4.005 – 5.005	132	862	0.862
5.01 – 6.00	5.005 – 6.005	49	911	0.911
6.01 – 7.00	6.005 – 7.005	72	983	0.983
7.01 – 8.00	7.005 – 8.005	17	1000	1
Mean Paths disrupted = 0.031376533			SD = 0.016855781	

Fig. 10 shows that the distribution of paths disrupted for 4 km radius is approximately symmetric. Majority of the paths disrupted were below 10.505% which comprises to 59.6%, while only 5.7% of the paths disrupted were above 18.005%. The mean paths disrupted is 0.094153646 with a standard deviation of 0.04817. Moreover, 77.3% of the paths disrupted ranged from 5% to 21.39%.
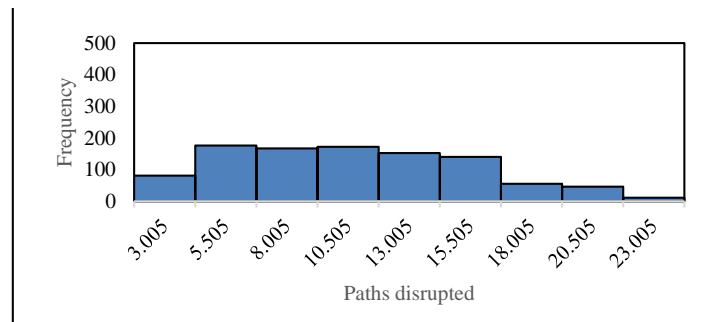

Fig. 10. Histogram of Paths disrupted for 4 km radius

Fig. 11 shows a comparison for the max paths disrupted. At 1 km radius, the max paths disrupted is at 4.86%. As such, 5% was chosen to act as the "elbow" of the graph and to be the critical point in the study.

Fig. 11 Comparison of max paths disrupted

Table 8 compares the mean RI with the mean Beta and Gamma indices of the Metro Manila road network. It shows that as the radius of destruction increases, the mean RI decreases along with the mean Beta and Gamma indices.

Table 8. Comparison of mean RI with mean Beta and Gamma indices.

Index	Gamma	Beta	RI
0.25 km	0.351906	1.05568	0.998158
0.5 km	0.351588	1.054727	0.996212
1 km	0.350597	1.051754	0.989858
2 km	0.347019	1.041019	0.968623
4 km	0.335965	1.007858	0.905846

5. Conclusion

Results of this study show that the Metro Manila road network is fairly robust up to 1 km radius of destruction, where a maximum of 4.86%% paths disrupted is incurred. However, the damage sharply increases for larger radii of destruction. 5% or more loss in the network's connectivity is noted in 13.8% of the nodes for the 2 km radius and in 77.3% of the nodes for the 4 km radius.

Acknowledgment

References
[1]

[14] Bell, M.G.H. and Iida, Y., 1997. *Transportation network analysis.* Chichester, New York, USA: J. Wiley.

[15] Immers, L.H., Stada, J.E., Yperman, I. and Bleukx, A., 2004. Robustness and resilience of transportation networks. *In*: Proceedings of the 9th International Scientific Conference MOBILITA. Bratislava, Slovenia.

[16] Santos, B. F., Antunes, A. P., & Miller, E. J. (2010). Interurban road network planning model with accessibility and robustness objectives. *Transportation Planning and Technology*, *33*(3), 297-313.

[17] Berdica, K., 2002. An introduction to road vulnerability: What has been done, is done and should be done. *Transport Policy,* 9 (2)**,** 117-127.

[18] Warfield, J. N. (1973). Binary matrices in system modeling. *Systems, Man and Cybernetics, IEEE Transactions on*, (5), 441-449.

[19] Harary, F., Norman, R. Z., & Cartwright, D. (1965). Structural models: an introduction to the theory of directed graphs.

[20] Open Street Map Wiki. http://wiki.openstreetmap.org/

[21] Metro Extracts. https://mapzen.com/metro-extracts/

[22] Isla, L. T. (2014). Osm-converter. https://bitbucket.org/droxisla/osm-converter/overview

[23] Open Street Map Lanes. http://wiki.openstreetmap.org/wiki/Key:lanes

[24] Open Street Map Highways. http://wiki.openstreetmap.org/wiki/Key:highway

[25] Teknomo, K. and Fernandez, P. (2012). A theoretical foundation for the relationship between generalized origin destination matrix and flow matrix based on ordinal graph trajectories. Journal of Advanced Transportation.