Sao Paulo Cycling Network Development Design
A Minimum Spanning Tree Approach

Zili, W.

DOI
10.1061/9780784483053.246

Publication date
2020

Document Version
Accepted author manuscript

Published in
CICTP 2020

Citation (APA)
Zili, W. (2020). Sao Paulo Cycling Network Development Design: A Minimum Spanning Tree Approach. In H. Wei, H. Wang, L. Zhang, Y. An, & X. Zhao (Eds.), *CICTP 2020: Transportation Evolution Impacting Future Mobility - Selected Papers from the 20th COTA International Conference of Transportation Professionals* (pp. 2916-2929). (CICTP 2020: Transportation Evolution Impacting Future Mobility - Selected Papers from the 20th COTA International Conference of Transportation Professionals). American Society of Civil Engineers (ASCE). https://doi.org/10.1061/9780784483053.246

São Paulo Cycling Network Development Design:
A Minimum Spanning Tree Approach

Zili W. Author, M.S.[1]

[1]Department of Civil Engineering, Delft University of Technology, P.O. Box 5048, Delft, 2628 CN; PH (031) 06-44785460; FAX (031) 01-52783179; e-mail: z.wang-15@student.tudelft.nl

ABSTRACT

Cycling is a heated topic in social media and a political hotspot in São Paulo. The implementation of bicycles took place in the city after the cooperation agreement signed by the Municipal Bureau of International and the Institute for Transportation & Development Policy in 2009. The recent 10-year development of the cycling infrastructure resulted in an unconnected and scattered network throughout the city. To improve the accessibility and increase the service coverage, the study proposed the minimum spanning tree to design a well-connected cycling network. A case study of 4 center districts has been researched. The new plan aims to serve 94.49% inhabitants within 350 meters of the walking distance and create links to the daily trip destinations in the regions, such as public transport stations, schools, shopping malls, hospitals, etc.

INTRODUCTION

São Paulo traffic is highly dependent on cars and public transport. However, it has been suffering from a lot of traffic congestion, resulting in large additional travel time for vehicles, and the public transport infrastructure is always saturated. The bicycle was proposed as an important mode of transportation by the cooperation agreement signed in 2009 by the Municipal Bureau of International and the Institute for Transportation & Development Policy and the Strategic Masterplan in 2014 of the Municipality of São Paulo. It meant the that establishment and improvement of the cycling infrastructure was adopted in a political view. But due to the regular change of the city mayor, the progress was slowed down tremendously. Thus, after the 10-year development of the bicycle infrastructure, São Paulo has an unconnected and scattered cycling network, and the share of cycling use in 2018 was 1% (Dixon et al., 2018). Recently, there arose a higher preference to use bicycles contributed by the heated discussion in the media and the simulation of the governmental policies that advocate that the transport mode is good for leisure and tourism (Medeiros & Duarte, 2013). The new upsurge in travelling on the bicycles is ongoing together with the development of the shared-bicycle market. The current bikeways in São Paulo have the problem of bad connectivity and small service coverage, which indirectly caused the high casualties with cyclists. Therefore, the research aims to develop an approach of designing a well accessible, safe, and comfortable bicycle network for São Paulo.

To achieve this goal, this study develops a new cycling network design model to find the optimal layout of bikeway networks in São Paulo urban areas. Minimum spanning tree is the core of the model. The proposed model is applied to bikeway planning in four center districts of the city (Butantã, Pinheiros, Alto de Pinheiros, and Jaguaré). It is also applicable to the other districts or cities in Brazil as long as the design goal is the same.

The paper structure is organized as follows. ‘‘Modeling concept’’ chapter describes the framework of the cycling network design model. “Data input” chapter introduces the input data. “Selection of candidate links” chapter selects the safe and comfortable paths as the candidate links out of the present road map. “Selection of demand nodes” chapter elaborates on the selection criterion and results of the origin and arrival locations of the cycling trips. “Formulation of link utility” chapter formulates the utility function and computes the utility of all the candidate links. “Route selection model” chapter uses the greedy algorithm for the cycling network design problem. ‘‘Model application’’ chapter applies the model to four districts in São Paulo and results in the final design of the cycling network. Finally, the ‘‘Conclusion’’ chapter summarizes the research.

MODELING CONCEPT

São Paulo has a highly-dense road system, meaning a lack of space for constructing new paths in the urban environment. The bikeways in the design have to attach to the present road network. In other words, the cycling network is the subset of the present road network. To some extent, this makes the network design problem simple. The key to the design model is selecting the links with the high utility from a set of the current roadways, and the collection of the selected objects is the result. Accordingly, the cycling network design problem in this study is an optimization problem essentially. The classic approach to solving the kind of problems is applied, as like setting up the objective(s) and the constraint(s) and then finding the optimal answer.

The complexity of the problem is derived from the multiple design requirements. In line with the research goal, the main objective is to maximize service coverage, and the sub-objectives are to minimize cyclist risk and to maximize cyclist comfort. Additionally, the new design should perfectly connect with the existing cycling infrastructure. To meet the multiple requirements, the model is designed as a primary selection step followed by an optimization approach. In the selection, this work achieves the sub-objectives and takes the previous cycling ways into account. It results in two sets of the links: one is the candidate link which has the small risks for cyclists and the acceptable riding comfort that are determined by the pavement types and the traffic intensities, and the elevations respectively. The other is the existing cycling ways that are indicated as the high utility in order to be include in the optimal result. In the optimization, the model finds the optimal cycling network out of the selection result, with the main objective to maximize service coverage and the constraints of the departure and arrival nodes and the current bikeways. The variables of the nodes and the link utility are elaborated on and computed in two dependent sections in the optimization approach. All the other variables of the pavement types, the traffic intensities, the elevations, and the present bikeways are accessed directly from the input data and maps. To conclude, the model framework contains four parts in order as shown in Fig. 1: data input, pilot selection approach, optimization approach, and the cycling network as a result. The model of the optimization is made up of three parts: the demand nodes, computing the link utility, and running the route selection model.

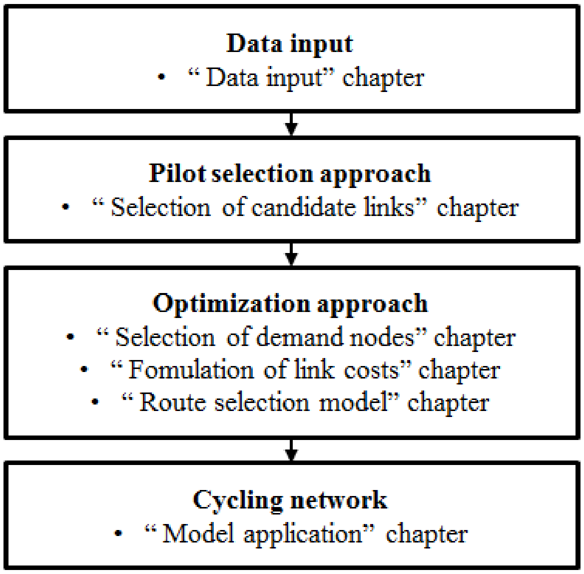


Figure 1. Model framework

DATA INPUT

The study relies on two data sources: GeoSampa Mapa (Prefeitura de Sao Paulo, 2019) and Mobility Monitoring Survey (CET, 2017). The input data and the variables as the derivation thereof are listed in Table 1.

Table 1. The input data and the derivation variables

Data	Name	Type	Source	Derivation variables
Geography	Declividade	shapefile	GeoSampa Mapa	Elevation
Demographic density	Densidade demográfica	shapefile	GeoSampa Mapa	Nodes, link utility
District	Distrito	shapefile	GeoSampa Mapa	Nodes, link sets
Road map	Logradouro	shapefile	GeoSampa Mapa	Road types, link sets
Registered entity	Uso Predominante do Solo	shapefile	GeoSampa Mapa	Nodes
Metro station	Metrô estacão	shapefile	GeoSampa Mapa	Nodes
Bus terminal	Terminal de ônibus	shapefile	GeoSampa Mapa	Nodes
Tram station	Trem estacão	shapefile	GeoSampa Mapa	Nodes
Bikeway	Rede ciclovaria	shapefile	GeoSampa Mapa	Present bikeways
Traffic volume and speed	Volume e Velocidade	number	CET	Road types, traffic intensities

SELECTION OF CADIDATE LINKS

To set up a set of roads where the cycling paths are appropriately attached, the candidate links are selected by three criterion: (1) the road elevation is suitable for riding a bicycle; (2) for the safety of the cyclists, the intensities and the speeds on the pavement should not be too high; (3) if there is the existence of the bikeways, the road should be in the set. Accordingly, the attributes of the road elevation, the road hierarchy, and the presence of the bicycle infrastructure are taken into account in the pilot selection approach. The following three paragraphs give the elaboration and the necessary data processing of the three attributes respectively.

São Paulo is a city with height variations. Steep roads make cycle riding difficult. Roads with a steepness of more than 5% are considered as not comfortable for riding a bike (Lin & Yu, 2012). Theoretically, the road gradient is determined by both the elevation and the direction of the road. That is to say, one road section with a longitudinal incline has two road gradients, of which the positive is for the uphill direction and the negative is for the downhill. It should be noticed here that even the steep road saves energy to ride in the down direction. It is dangerous to cycle on the dramatic downhills, and in this sense, they are not applicable to cycling. To achieve comfortable riding, only the links with a slope of the absolute value of less than 5% are selected. The study classifies the gradients into four categories: the absolute value less than 5%, between 5% and 25%, between 25% and 60%, and more than 60%. The input data in the last chapter is the slope per area of 100 square meters. It is necessary to process the data to get the slopes of the roadways. The data is processed with the assumption that the slope of an area is used as the slope of the roads in the area.

At present, it is fast and heavy motor traffic that threatens cyclists' lives (Jacobsen & Rutter, 2012). To aid the jurisdiction in planning for safe cycling infrastructure, selecting the appropriate bikeway types for the current network is required. The current road system in the city contains 49 pavement type that are determined by the road geometry and the road functions (Prefeitura de Sao Paulo, 2019). Some categories includeing the highways and the fast transit roads only serve the motor vehicles, and non-motor vehicles are prohibited (Brazil, 1997). Some that carry low volumes of traffic at low speeds, and others have high traffic flows and driving speeds (CET, 2017). Generally, there are three types of bikeways that are applicable to different traffic conditions: the bicycle paths, bicycle lanes, and shared lanes. In this study, the bicycle path is defined as the bikeway only dedicated to cycling and physically separated from motorized traffic. The bicycle lane is the two-way facility that attaches to the edge of the motor-traffic path typically carrying bicycle traffic. The shared lane is the facilities used by the mixed traffic of non-motorized and motorized vehicles. Rural roads with good sight distance, serving low volumes at the speeds at most 89 km/h provides an enjoyable and comfortable bicycle experience with no need for bike lanes or paths (AASHTO, 2012). The condition should be more critical in the urban area, because the traffic is denser usually and the sights are affected by the complex surroundings to common sense. Due to the characteristics of bicycle lanes and paths, the lanes that have the possibility of sharing the facilities with the motor vehicles are applicable to the minor arterial ways, but not safe for the major arterial roads that commonly have more traffic volumes and higher speeds. The cycling path can provide the suitable infrastructure for cyclists on major arterial roads (AASHTO, 2012). Accordingly, the road hierarchy in Tab. 2 is made for distinguishing whether the road is safe to ride a bike and for determining the suitable bikeway types attached to it.

Table 2. Road Categorization

Road category	Characteristics	Bikeway selection
Access road	Speed limit lower than 60 km/h	The share lane
Minor arterial road	Speed limit between 60 km/h and 110 km/h, peak volumes less than 2000 veh/h	The bicycle lane
Major arterial road	Speed limit between 60 km/h and 110 km/h, peak volumes more than 2000 veh/h	The bicycle path
Freeway	Speed limit higher than 110 km/h	No suitable cycling facility

The present cycling infrastructure is unconnected and scattered throughout the city. The core goal of the research is to develop the network on the basis of the existing bikeways. The cycling map is input from GeoSampa (Prefeitura de Sao Paulo, 2019), and these links are selected in the set of the candidate links.

SELECTION OF DEMAND NODES

To design the cycling network, one necessary approach is to determine the points where the network should connect. The core of the step is to figure out where the large number of cycling trips originate and arrive. Due to less usage of bicycles in São Paulo (Dixon et al., 2018), we lack the data of the departures and the destinations of the bicycle traffic. Thus, this study assumes that the places where a large number of pedestrians flow from or to are the potential origins or arrivals of the cycling trips, which will be connected in the new cycling network. Generally, the surroundings and the insides of public transport terminals, residential districts, office skyscrapers, large shopping malls, and public service organizations (hospitals, schools, etc.) have large flows of pedestrians. All the landmarks are accessible directly from GeoSampa (Prefeitura de Sao Paulo, 2019) as the form of the node maps in GIS. The residential districts are marked in the population map as the division of São Paulo into 18953 zones, each of which has a number of the registered residents in the zone (Prefeitura de Sao Paulo, 2019). Thus, the data is processed by translating the population map to the node map that contains the population aggregation points. It should be noted here that not all the population aggregation points are included, because it is neither practical nor feasible to compute a network connecting the large set of nodes. There should be a node selection, which is a typical maximum coverage problem. The method of the data processing has two steps: (1) find the centroids of all the areas and convert the population of the areas to the nodes; (2) select the important population aggregation points by the Maximum Coverage Location Model (Daskin, M. S., 1983). The model is applicable to the case because it is feasible to reduce the nodes and results in the demand nodes with the maximum service coverage. To conclude, all the locations of public transport terminals, office skyscrapers, large shopping malls, and public service organizations (hospitals, schools, etc.) are selected as the demand nodes in the case, as well as the important population aggregation points representing the residential districts.

FORMULATION OF LINK UTILITY

The research aims to develop the cycling network with the aim of maximizing the service coverage. Therefore, the cycling network is made up of two parts: one is the existing bikeways, and the other is selected out of the candidate links according to the link utility. In the study, the final design includes the present cycling facilities by assuming the high utility on those links. The link utility of the candidate link is defined as the number of people the link can serve with the consideration of the preference to ride on the different types of roads. The number of served persons is counted up based on the population map. The preference of various road types is on the basis of the previous literature on cyclist behavior.

The service coverage of the road is defined as the number of inhabitants who are able to reach the cycling network within an acceptable walking distance with their bicycles. People are willing to travel without being on the dedicated bicycle infrastructure for around four blocks (Lin & Yu, 2012), in São Paulo, that is about 350 meters. Accordingly, the coverage is formulated as the population in the area around the link with an offset of 350 meters. Though some inhabitants are counted in several adjacent links, the overlap does not cause the problem of the network design, because the selection by the route selection model of any link is independent. But it affects the final result of the total number of people served. The sum of the coverage of the links in the network design is more than the actual service coverage because of the overlap. The accurate computation is establishing the zone around the network with an offset of 350 meters and adding up the population in the area.

Apart from the service coverage, the preference to riding on the various roadways is taken into account in the link utility. According to the survey of the stated preference of 695 commuting cyclists, the routes on minor and major arterials (relative to routes on residential roads) are less likely to be chosen for commuting purposes (Dey et al., 2018). Some findings and the derivations thereof of the research are in Tab. 3. The values of probability will be used in the function of the link utility.

Table 3. Choice behavior of cyclists (Dey et al., 2018)

Road types	Coefficient	Probability of chosen
Access road (Residential road, in the Dey's paper)	1	74.16%
Minor arterial road	-0.398	18.33%
Major arterial road	-1.29	7.51%

The study formulates the link utility as Equation (1), where U is the link utility; P is the probability of the road is chosen to ride a bicycle; C is the number of inhabitants who lives in the area within 350 meters far from the link. If the link has cycling facilities, the utility is assumed to add 99999 for making it a high value, in order to ensure the link is selected by the route selection model.

$$U = P * C + E \qquad (1)$$

$$\mathbf{E} = \begin{cases} 99999 & \text{if the link has the bikeway} \\ 0 & \text{if the link has no bikeway} \end{cases}$$

ROUTE SELECTION MODEL

The route selection model is to find the optimal routes in the whole network between the departure locations to the destinations as determined according to the utility of links. The optimization approach is a greedy algorithm of which the classic paradigm intends to a global or local optimum by following the problem-solving heuristic and making the locally optimal choice at each step (Cormen et al., 2001). In terms of the travelling route problem, the greedy strategy follows the approach at each step of the journey and visits the nearest unconnected nodes in the next step until the heuristic terminates in a countable number of steps or all the nodes that should be connected had been reached.

Minimum spanning tree (MST), as one greedy algorithm, has been widely-applied in the network design in multiple fields, but rather new to the cycling network design. It is always a dilemma that lies in solving the time and complexity of a scenario that the model is designed to capture. The selection of MST is a tradeoff, which is efficient to solve the magnitude of the network fast and takes as many of features as the design requires into account. The method in the research is to select the cycling routes through the demand points from the candidate links. The algorithm is chosen because it is well-developed and broadly applied to many transportation network design problems. The study applies TransCAD to run MST model (Kruskal, 1956). It has the classic result that is a subset of the candidate network of the shortest total distance, undirected and connecting all the nodes without any cycles. The computation process runs in $\log_{\mathrm{M}}\mathrm{N}$ times, where N and M denote, respectively, the number of the connection components and the candidate links of the case. The algorithm starts at any random point and then goes throughout the whole network, and for each step, it selects the link of the shortest length with no loops. In the study, the link attribute applied to the algorithm is the reciprocal of the link utility, rather than the length, so that the final spanning tree is the result of the maximum of the total utility connecting all the demand nodes.

MODEL APPLICATION

The design model is applied to four center districts of São Paulo (Butantã, Pinheiros, Alto de Pinheiros, and Jaguaré), in Fig. 2. The input data mainly contains the roadmap (Fig. 3a), the population map (Fig. 3b), the elevation map (Fig. 3c), the public transport map (Fig. 3d), the present cycling map (Fig. 3e), the land use map (Fig. 3f), and traffic volumes and speeds (CET, 2017). The properties in the maps (Fig. 3) are explained in GeoSampa (Prefeitura de Sao Paulo, 2019).

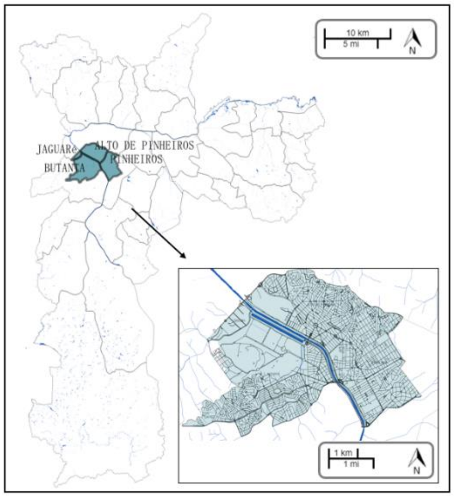


Figure 2. São Paulo and the case study area

The total road map in the study area contains 7597 links, 57.47% of which are access roads, 9.73% for minor arterial roads, 31.21% for major arterial roads, and 1.59% for freeways, as shown in Fig. 4(a). The pilot selection approach is modelling in TransCAD. In the step, 26.46% of the total number of links in the study area are filtered out including 2249 steep links, mainly located in the northeast and southwest regions of the study area in Fig. 4(b), and 121 links representing the freeways of Av. Eng. Billings and Marginal Pinheiros in Fig. 4(a). All the other links are in the set of the candidate links, which has 5587 links in total. The set of the present cycling facility in Fig. 4(c) has 1116 links, where 11.4% of the inhabitants in the study area are within a 350-meter walking distance. It should be noticed that there is a fast cycling path constructed along Marginal Pinheiros adjacent to the river. In the road map, the path is aggregated with Marginal Pinheiros. Thus, the freeway is excluded firstly but then added in the present cycling facility set.

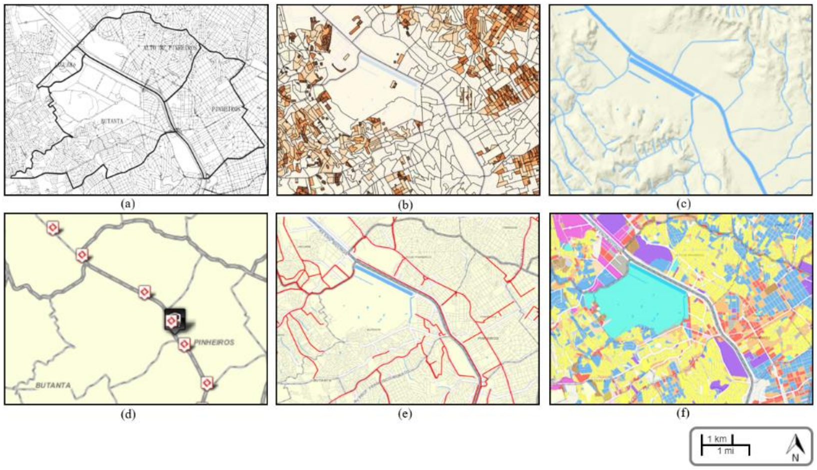
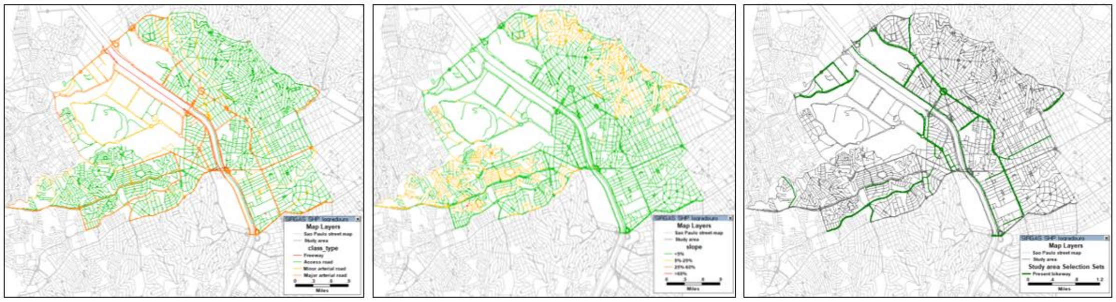


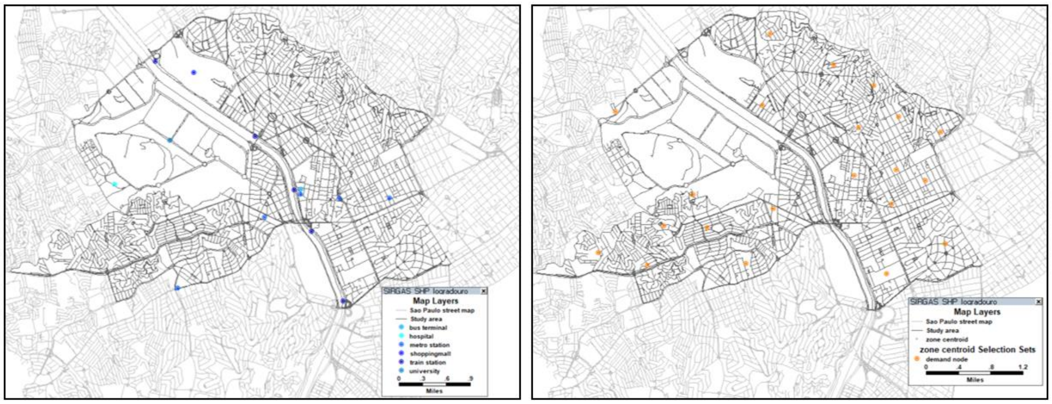
Figure 3. Input maps:
(a) road; (b) population; (c) geography;
(d) public transport; (e) bikeway; (f) land use



(a) (b) (c)

Figure 4. Road classification by:
(a) hierarchy; (b) slope; (c) the existence of bikeways

After setting up the candidate link set, the optimization approach is modeled in TransCAD, including selecting the demand nodes, computing the link utility, and modeling minimum spanning tree in order. In the study area, there are 5 metro stations (Morumbi, Butantã, Pinheiros, Faria Lima, and Fradique Coutinho), 5 train stations (Villa Lobos, Cidade Universitària, Prinheiros, Estacão Hebraica, and Estacão Cidade Jardim), 1 bus terminal (Pinheiros), 1 large shopping mall (Shopping Villa Lobos), 1 university (Universidade de São Paulo), and 1 university hospital (Hospital Universitário de São Paulo, at the edge of campus), shown in Fig 5(a). Thus, 14 nodes representing the important buildings are selected in the demand-node set. In addition, among 409 centroid points converted by the populated zones, 21 nodes are selected by “Facility Location” module in TransCAD in Fig. 5(b). As a result, 35 nodes are in the set of the demand nodes which will be connected in the new cycling network, and any node of the total network can reach the nearest demand nodes within 300.88 meters on average at most 852.95 meters.

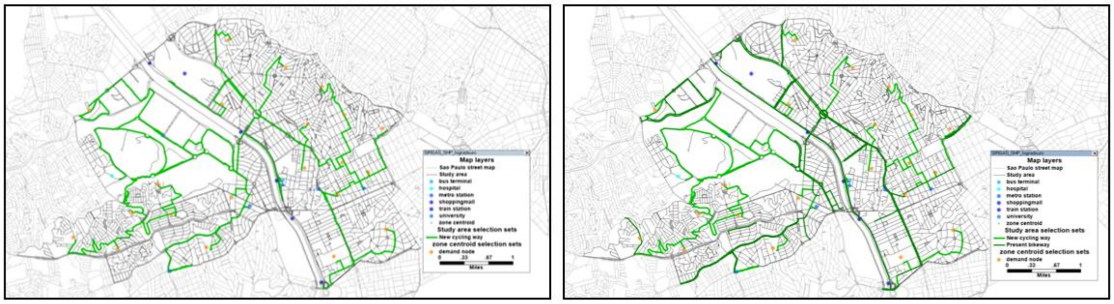


(a) (b)

Figure 5. Demand nodes:
(a) important facilities; (b) selection results of residential locations

The formulation of link utility is run by “Edit Attribute Table” module in QGIS. The results range from 6.42 to 112606.2. The value above 99999 indicates the existence of the bikeway on the link, and the utility of the candidate links is between 6.42 to 13370.31.

According to the link utility, the route selection model applied by “Minimum Spanning Tree” module in TransCAD finds the optimal result including 581 links, as shown in Fig. 6(a). These links extend the cycling network to the demand nodes and connect some independent bicycle lanes. They are composed of routes on 21 roads, including 12 access roads, 4 minor arterial roads, 4 major arterial roads and 1 freeway (the existing cycling facility). Aggregated with the present cycling ways, the final design of São Paulo cycling network in Fig. 6(b) accounts for 18.94% of the whole road infrastructure. The new cycling facilities constitute 40.37% of the cycling network, and the others are the present bikeways. In the new cycling facilities, 44% are the access roads, while the arterial roads account for 50% and the others for 6%. It is worthy of noticing that in the current bikeways, the proportion of the arterial roads as 61% is larger than the access road as 35%. But the new design of the study improves the share of the access roads apparently. The result is logical, because in the context that São Paulo has developed some cycling ways, the access roads will be potentially chosen to reach the destinations for the last mile.



(a) (b)

Figure 6. Cycling network design:
(a) the optimal result of MST; (b) the aggregation of the present and new bikeways

In this study, it is important to produce the statistics of the inhabitants who can be served by the new cycling network. The service coverage is defined above as the region around the network with an offset of 350 meters. It is drawn by “Buffer” module in QGIS, as shown in Fig. 7(a). 388 zones are in the service, and 21 are not, as shown in Fig. 7(b). According to the statistics results, among the total population in the study area of 168,236 people, 94.49% are able to reach the new bicycle infrastructure within the walking distance of 350 meters. Therefore, the design aim of maximizing service coverage is achieved ideally.

(a) (b)

Figure 7. Service coverage of the new cycling network
(a) service zone; (b) centroids of the served and unserved zones

CONCLUSION

This research is to develop the bicycle network in São Paulo to improve cycling use. Through the investigation of the present traffic of the city, the biggest problem concerning the existing bikeway turns out to be the connectivity and accessibility, which is absent on many zones and cannot be accessed most of the daily trip destinations such as public transport stops, schools, shopping malls, etc. The study designs a model to extend the present cycling ways to the potential demand origins and uses the minimum spanning tree model to achieve the maximum service coverage. Additionally, the model takes riding comfort and safety into account. The model is applied to four center districts of São Paulo as the case study. The new bicycle infrastructure is designed in the study area with the availability that 94.49% of the inhabitants are capable to access the network within an acceptable walking distance of 350 meters. The report shows the connected cycling network design in Fig. 6. Although this network is created for only a pilot area of São Paulo, the model is designed to be implemented in other areas of the city. To conclude, the study provides a solution to the cycling network design problem in the context of São Paulo.

REFERENCES

Dixon, S., Irshad, H., Pankratz, D. M., and Bornstein, J. (2018). “The deloitte city mobility index.” *Deloitte Insights*, <https://www2.deloitte.com/content/dam/Deloitte/za/Documents/Consumer_Industrial_Products/ZA_Deloitte_Global_Mobility_Index_October_2018.pdf> (Oct. 24, 2019)

Medeiros, R. M. and Duarte, F. (2013). “Policy to promote bicycle use or bicycle to promote politicians: bicycles in the imagery of urban mobility in Brazil.” *Urban, Planning and Transport Research*, 1(1), 28– 39.

Prefeitura de Sao Paulo (2019). GeoSampa Mapa Retrieved 2019-02-19, from http://geosampa.prefeitura.sp.gov.br

CET (2017). *Pesquisa de Monitoração da Mobilidade : Mobilidade no sistema viário principal volume e velocidade - 2017 (in Portuguese)*

Lin, J. J., and Yu, C. J. (2012). “A bikeway network design model for urban areas. ” *Transportation*, 40(1), 45–68.

AASHTO (2012). *Guide for the development of bicycle facilities fourth edition.*

Daskin, M. S. (1983). “A maximum expected covering location model: formulation, properties and heuristic solution.” *Transportation science*, 17(1), 48-70.

Dey, B. K., Anowar, S., Eluru, N., and Hatzopoulou,M. (2018). “Accommodating exogenous variable and decision rule heterogeneity in discrete choice models: Application to bicyclist route choice. ” *PloS one*, 13(11), e0208309.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2001). *Introduction to algorithms second edition.* The Knuth-Morris-Pratt Algorithm.

Kruskal, J. B. (1956). “On the shortest spanning subtree of a graph and the traveling salesman problem. ” *Proceedings of the American Mathematical society*, 7(1), 48–50.

Brazil. (1997). Lei nº 9.503, de 23 de setembro de 1997. Institui o Código de Trânsito Brasileiro. Diário Oficial da República Federativa do Brasil, Brasília, 24 set. 1997. Available at: <http://www.planalto.gov.br/ccivil_03/leis/l9503.htm> (in Portuguese)

Jacobsen, P., and Rutter, H. (2012). “Cycling safety.” Chapter 7 in: *City Cycling.* MIT Press, Cambridge, MA, 141-156. ISBN 9780262517812.