International Journal of Sustainable Transportation

ISSN: 1556-8318 (Print) 1556-8334 (Online) Journal homepage: http://www.tandfonline.com/loi/ujst20

Data-mining, gis and multicriteria analysis in a comprehensive method for bicycle network planning and design

Ms Thais De CÁssia Martinelli Guerreiro, Dr Janice Kirner Providelo, Dr Cira Souza Pitombo, Dr Rui Antonio Rodrigues Ramos & Dr Nelson Rodrigues Da Silva Antonio

To cite this article: Ms Thais De CÁssia Martinelli Guerreiro, Dr Janice Kirner Providelo, Dr Cira Souza Pitombo, Dr Rui Antonio Rodrigues Ramos & Dr Nelson Rodrigues Da Silva Antonio (2017): Data-mining, gis and multicriteria analysis in a comprehensive method for bicycle network planning and design, International Journal of Sustainable Transportation, DOI: 10.1080/15568318.2017.1342156

To link to this article: http://dx.doi.org/10.1080/15568318.2017.1342156

Accepted author version posted online: 21 Jun 2017.

Journal Title: International Journal of Sustainable Transportation

DATA-MINING, GIS AND MULTICRITERIA ANALYSIS IN A COMPREHENSIVE METHOD FOR BICYCLE NETWORK PLANNING AND DESIGN

Ms THAIS DE CÁSSIA MARTINELLI GUERREIRO[1], Dr JANICE KIRNER PROVIDELO[1], Dr CIRA SOUZA PITOMBO[1], Dr RUI ANTONIO RODRIGUES RAMOS[2], Dr ANTONIO NELSON RODRIGUES DA SILVA PhD[1]

[1]UNIVERSITY OF SAO PAULO, Sao Carlos School of Engineering - Dep. of Transportation, Av. Trabalhador Sao-carlense 400, Sao Carlos, 13566-590 Brazil

[2]Universidade do Minho, Department of Civil Engineering, Campus de Gualtar, Braga, 4710-057 Portugal

Corresponding Author: Email: jkirner@gmail.com

Email: guerreiro_ec@yahoo.com.br

Email: cirapitombo@sc.usp.br

Email: rui.ramos@civil.uminho.pt

Email: anelson@sc.usp.br

Email: anelson@sc.usp.br

ABSTRACT

In many Brazilian cities, the most common procedure for planning cycling networks is using aggregated population data in census tracts, which may not take into account the true origin and destination of trips. It may also not identify potential users of a particular mode of transport. This is particularly important considering that implementing cycling infrastructures should be based on the assumption that they are able to meet the users' needs. Therefore, the aim of this study is to develop and adopt an objective method to design and compare cycling networks based on data-mining of disaggregated origin-destination data, GIS resources and multicriteria analysis techniques. The method follows three steps: a) identifying potential users based on real user profiles, b) designing proposed cycling networks and c) a comparison between the networks proposed in this study and those developed by the municipality selected as a case study, considering real and potential users, as well as cost and benefit criteria. As a positive outcome, using disaggregated data allows for a reasonable estimate of the number of people served by the networks, a detailed analysis of their proximity to the infrastructure, as well as identifying potential users. Comparing cycling networks considering cost and benefit criteria shows that the chosen criteria were effective. It was also determined that the cycling network of the studied city poorly serves bicycle transport users, if compared to the proposed networks. These findings indicate that appropriate methods for planning cycling networks are still needed.

Keywords

cycling infrastructure, bicycle network planning, data-mining, GIS, multi-criteria analysis, potential cyclists

1 INTRODUCTION

After many years of investing in road infrastructure for automobiles coupled with the lack of urban planning and not adopting sustainable mobility alternatives, the situation in many cities in developing countries is almost in a state of collapse. As a result, there has been a rise in air pollution levels, high levels of consumption of natural resources, segregation of urban spaces, an increase in the number of traffic accidents and more urban immobility (Benicchio, 2012).

One of the alternatives to avoid this worsening state of affairs is by planning cycling networks. In most Brazilian cities, the most used practical procedure for planning new cycling networks is still based on exploratory and incremental designs. In many situations, this procedure is inefficient as it considers population data associated with the official census tracts. The main point of this procedure is that population data obtained in this way are extremely aggregated, which causes two problems. The first one is the data aggregation bias known as "ecological fallacy", which comes from the false assumption that the population is homogeneously distributed within the census tracts. The second problem is the fact that people who cycle, either current users or potential ones, are not clearly identified in this population. Current users can be defined as those who already use bicycles for daily commuting. Potential users, on the other hand, are those that currently use other modes of transport, but could switch to cycling in the future.

Specifically regarding potential users, they are most commonly identified by conducting stated or revealed preference surveys about the studied population. However, the survey answers are not always consistent with the actual behavior of the respondents, either because the responses are not necessarily true or because the respondents decided to act differently of what was said

ACCEPTED MANUSCRIPT

when they were questioned about the matter. In addition, these surveys can be costly as interviewers need to be trained and considerable time for data collection and processing information is required. In many cases, such as for smaller municipalities with limited resources, these conditions can make these surveys completely unfeasible.

However, planning a cycling network is not so simple. It is necessary to anticipate if the networks (existing or proposed) can effectively serve real and potential users. A poorly planned network that users do not find interesting is a waste of public resources and is, therefore, an inefficient infrastructure. In this case, there is even a likelihood that the infrastructure could increase the number of accidents involving cyclists as the users will search for alternatives in shared traffic to complete the trip at both ends (i.e. from the origin and to the destination). One way of measuring the effectiveness of the cycling network is, for example, by using a multicriteria analysis of the costs and benefits of the considered networks.

In this context, this study aims to answer the following questions related to planning cycling networks:

- How can potential users be identified without using stated or revealed preference surveys?
- How can we plan cycling networks using disaggregated population data?
- Which criteria must be considered to evaluate proposed cycling networks and how can we compare them with the existing infrastructure?

In order to answer these questions, an objective method to design and compare cycling networks was developed and used, considering real and potential users. The method is considered objective due to the fact that actual travel data is used and also because the listed costs and benefits are tangible and can be objectively calculated. On the other hand, the method can also be considered subjective as there is a certain amount of subjectivity when choosing the variables considered important to determine the potential users and to compare the networks. This is not considered a problem, however, as the proposed method answers questions related to planning cycle networks.

The method used disaggregated origin-destination data. This can be described as data containing the exact geographical location (latitude and longitude) of the origin and destination of each trip made by real users, which were then associated to data-mining techniques (to identify potential users), Geographic Information System (GIS) resources (for the cycling network planning) and multicriteria analysis techniques (for an effective comparison of the studied networks). The method was developed and adopted in a Brazilian medium-sized city that has an incipient cycling infrastructure.

2 LITERATURE REVIEW

Based on the literature review, four main topics regarding this study can be highlighted. These topics are addressed in this section and can be categorized into: i) cycling networks in Latin America, ii) identifying potential users of cycling transport, iii) cycling network design, and iv) using GIS associated with multicriteria analysis.

2.1 Cycling networks in Latin America

Nowadays, various municipalities in Latin American countries are recognizing the importance of cycling networks to reduce urban mobility problems and environmental issues by implementing urban policies concerning cycling infrastructures. Some examples are plans for the cities of

ACCEPTED MANUSCRIPT

Bogotá, Colombia (IDU, 2014), Santiago, Chile (Plataforma Urbana, 2015) and San Pedro Garza Garcia, Mexico (Implan, 2012).

Brazil, in particular, has never had a comprehensive cycle plan at a national level. Since the 1950s, riding bicycles has been linked to sports and leisure activities, often also associated to children. In this scenario, car use and ownership is associated to a high social status, while bicycles hold a second-class position in social ranking (Medeiros and Duarte, 2013).

In addition to the social issues involved, the tax burden applied to the purchase price of a new bicycle, compared to a car, can also be a disincentive to cycling. Currently, while the embedded tax on a new automobile price is around 30%, for a bicycle this percentage exceeds 70% (Vá de Bike, 2014). Some actions, still incipient in the legislative field, began to be developed in an attempt to reduce the price of bicycles, as well as increasingly encourage their use for urban trips, such as in the case of the following bills. Bill 4997/2013, which deals with taxes on industrialized products, would exempt bicycles from taxes. The aim of bill 6474/2009 is to set up a bicycle program in Brazil (Programa Bicicleta Brasil in Portuguese) to encourage people to use bicycles aiming at improving urban mobility conditions.

Two bills amend Federal Law No 7.418 dated 16 December, 1985, which created the Transport Voucher: bill 6418/2013 and bill 6724/2013. It is stated in the first one that: “Granting the benefit hereby implies: (…) II - the equivalent cash payment of the Transport Voucher to the worker, pursuant to item I of this article, when he/she chooses to commute to work by bicycle.” Thus, the Transport Voucher will not necessarily be used to pay for the bicycle (or its purchase), but rather for using it. In the case of bill 6724/2013, employees would be allowed to use travel vouchers to pay for bicycles.

In addition to the bills already mentioned, the only Federal law that is related to this subject is Law number 10.257/2001 (known as the City Statute), which requires cities with over 20,000 inhabitants to develop Mobility Plans. Compliance to this law can already be seen in some Brazilian cities, especially state capitals with more than 500,000 inhabitants, such as Brasilia (the capital of Brazil), Porto Alegre, Rio de Janeiro (host of the 2016 Olympics), Curitiba, Belo Horizonte, Aracaju, among others (PDTU, 2010; Ferrari, 2008; SMAC, 2010; IPPUC, 2013; BHTRANS, 2012; Cardoso, 2015, respectively).

In São Paulo, the largest Brazilian city, the Mobility Plan states that by 2030 more than 1,600 km of cycle paths and cycle lanes will be built, as well as other cycling infrastructures around the city (Prefeitura do Município de São Paulo, 2015). This breakthrough in the development of guidelines to build an extensive cycling infrastructure has been largely demanded by activist movements since 2002. In addition, these movements also require the commitment from political authorities to effectively take actions to encourage cycling. One example is the campaign for mayor of São Paulo for the term 2012-2016, in which the five candidates participated in actions involving cycling (BBC, 2015).

Currently, not only in São Paulo, but also in other Brazilian cities, many of the recently built cycling infrastructure segments are located in places with low bicycle travel demand, both real and potential. Furthermore, they are usually not connected to each other and result only in a political slogan of the current municipal administration, not contributing to the safety and agility of bicycle trips. One reason for this situation could be the lack of studies or guidelines to develop a mobility master plan considering cycling networks, as well as the lack of clear objective methods.

ACCEPTED MANUSCRIPT

2.2 Identifying potential users

Analysing questionnaires conducted with a sample of the study population is one of the most used strategies to identify potential bicycle users. The answers provided in these questionnaires often make it possible to identify the respondent's preferences, choices and decisions regarding bicycles and using them for daily trips, as discussed by Ortúzar et al. (2000), Mackett (2001), Bergström and Magnusson (2003), Dill and Voros (2007), Larsen et al. (2013), Wooliscroft and Ganglmair-Wooliscroft (2014), Wang et al. (2014) and Zhang et al. (2014).

As examples of factors that may interfere with using bicycles for commuting on a daily basis, Wuerzer and Mason (2015) concluded that not always the distance to be traveled to a university campus is the most important factor for students when deciding whether or not to use the bike every day. In some situations, cycle paths, cycle lanes and bicycle parking can be more efficient and attractive to stimulate riding a bicycle than the distance that needs to be covered to access the campus. According to Boussauw et al. (2014), short distances between home and school (primary), derived from a spatial distribution of schools adjusted to sustainable trips, can contribute to the increase in bicycle use on a daily basis. According to Castillo-Manzano et al. (2016), bicycle owners made longer trips than users of public shared bicycles as bicycle ownership offered greater independence, flexibility, comfort and convenience to adapt to the physical effort required. For Heinen et al. (2010), factors such as the built environment, moderate temperatures, reasonable distances, travel costs and travel times were considered positive for bicycle use in daily trips.

It is important to mention that although the respondents may have had positive attitudes towards using bicycles while the survey was being conducted, there is still no guarantee that he or she

 ACCEPTED MANUSCRIPT

will use this mode of transport in the future. This occurs because many decisions stated by the respondents are subject to change, as unforeseen situations may happen and individual skills can change over time.

Nevertheless, despite the increasing number of authors who are currently researching this topic, we were not able to find robust alternatives in the literature review to identify potential users. However, in the literature consulted, no alternatives are presented to identify potential users other than using questionnaires. Therefore, our aim is to attempt to fill this gap with an alternative based on using a data-mining technique applied to disaggregated population data.

2.3 Planning and designing cycling networks

Specifically regarding planning methods for cycling networks, some studies have been developed over the years based on different approaches. Many of them consider the use of a Geographic Information System (GIS) associated with population origin-destination data.

In the 1990s, Huang and Ye (1995) used a GIS to develop a road selection procedure based on origin-destination aggregated data and a gravity model to define cycling routes in the city of Berkeley, California. Aultman-Hall et al. (1997) used a GIS to determine the characteristics of 397 actual routes used by cyclists in Guelph (Ontario, Canada) and to compare them with the shortest path routes designed using a GIS software (*ArcInfo*) between each origin-destination pair. Considering the location of trip generators, Yamashita et al. (1998) proposed a methodology that integrated a Geographic Information System (GIS) with Remote Sensing (RS) resources to plan a cycling transportation system. More recently, Guerreiro and Rodrigues da Silva (2013) used a GIS for spatially locating disaggregated data obtained from an origin-destination survey, which were later used in the proposal of a cycling network to access a

 ACCEPTED MANUSCRIPT

university campus. Afterwards, Guerreiro et al. (2013) assessed the potential impacts of the proposed network on different types of campus users, as well as their current modes of transport (particularly walking and car trips). Segadilha (2014) combined a GPS (Global Positioning System) with a GIS for a comparative analysis of cycling trips in the city of São Carlos, in which actual routes were compared with the routes proposed by a transportation-specific GIS (*TransCAD*), considering the same points of origin and destination. A different approach was explored by Bergman e Oksanen (2016), who combined data of a mobile sports tracking application used, the platform Open Street Map and algorithms for providing automatic route suggestions for bicyclists. Another interesting contribution to the field was the Cycling Accessibility Index (CAI), which was developed by Saghapour et al. (2017) to measure cycling accessibility levels in terms of diversity of different land uses, number of activities in statistical areas, and the travel impedance between origins and destinations.

Lovelace et al. (2015) used aggregated population data in administrative districts, the number of trips between the districts, and desired lines between each origin-destination pair. The width of the desired line indicated the magnitude of the number of trips between each origin-destination pair. From the path covered between each pair, the authors were able to identify which roadways were used for the trips. Knowing the road gradient and the distance traveled, the authors developed a regression model to determine the number of trips made by bicycle, considering a few scenarios. More recently, Sousa and Sanches (2015) used responses from questionnaires answered by active cyclists who work in previously chosen trip generators to propose a generic procedure for defining cycling networks. In a more recent study, Boettge et al. (2017) developed an approach that surveys cyclists concerning level of stress along routes ridden. The mapped outcomes can be used by planners to identify high-stress routes as targets of new infrastructure or information to direct cyclists to safer routes. In general, the studies mentioned above used aggregated population data (from trip generators and administrative districts), which often do not represent the reality of trips. Thus, this research chose, as an alternative, to use disaggregated population data, which include the geographical information of the origin-destination pairs, associated with the shortest path algorithm in a GIS. By using this tool, the travel routes between the pairs could be determined, and, therefore, cycling networks were proposed for the city under study.

2.4 Using a GIS associated with multicriteria analysis

Although using a GIS is already consolidated for planning cycling networks, in many cases it should be combined with other techniques, such as Multicriteria Decision Analysis (MCDA) to seek for better results.

In recent years, various authors have used MCDA in studies related to cycling. Kirner (2006) proposed a method to define cycling routes in urban areas, considering travel demand, location of trip generators and bicycle level of service on roads with shared traffic as criteria for analysis. In Rybarczyk and Wu (2010), GIS and MCDA have been used to integrate supply and demand criteria for planning cycling infrastructures, considering main trip generators. Hsu and Lin (2011) proposed a planning procedure of a cycling network (using a shortest path algorithm considering traffic zones) and an evaluation of existing roadways using a Fuzzy Analytic Hierarchy Process. While Neri (2012) proposed and adopted a methodology to establish a potential cycling network for the city of Maringá, Brazil, Larsen et al. (2013) proposed a method to design a cycling infrastructure for the city of Montreal, Canada. The first study explored survey data and information from relevant urban features (topography, climate, shape and size of the city), main transport routes (identifying areas and cycling trip attraction/production hubs) and networks (connectedness of all considered roadways). The second study combined responses from a household origin-destination survey and a stated preference survey with criteria related to road segments. In the analysis of four different scenarios of cycling routes in the city of Athens, Greece, a GIS was used by Milakis and Athanasopoulos (2014) to locate eight trip generators, combined with evaluation criteria and weights obtained from groups of cyclists and researchers. Some of the more important evaluation criteria while analysing these different scenarios were: difficulty in riding a bicycle, density of crossings, traffic intensity and speed, legibility, natural environment, built environment, accessibility to activities, centrality, accessibility to urban parks and accessibility to metro/railway stations.

The literature review shows that none of the presented studies use multicriteria analysis techniques for existing cycling infrastructures, but only for developing new networks. Therefore, this study also intends to fill this gap by applying multicriteria analysis (considering costs and benefits) not only to cycling infrastructures at the design stage, but also to existing facilities.

3 METHODOLOGY

In this study, in order to meet the objective of developing and adopting an objective method to design and compare cycling networks based on disaggregated origin-destination data-mining, GIS resources and multicriteria analysis techniques, three steps are suggested: a) identifying potential users based on real user profiles, b) designing proposed cycling networks, and c) making a comparison between the proposed networks and the ones developed by the municipality, considering real and potential users, as well as cost and benefit criteria.

In the first step, which is to identify potential users, it is suggested using a data-mining technique applied to disaggregated data from an origin-destination survey including information on users' mode choice, as well as their socioeconomic data. Inserting data into a data mining package should be preceded by analysing the data consistency, choosing and classifying the information (variables) to be used and finding the best way to measure the data in the software. It is also recommended that the choice of variables is made according to their availability in the origin-destination survey and is based on their relation with bicycle use, such as those presented in the studies by Boussauw et al. (2014), Castillo-Manzano et al. (2016) and Heinen et al. (2010).

After choosing the dependent variable (mode choice), the independent variables and determining two travel samples (training and test samples), it is proposed to use a data-mining technique (for example, Decision Trees), which can identify real users' patterns to determine the potential users and their probability of cycling.

It is recommended that the second stage (designing the cycle networks) should be carried out based on the disaggregated data of the population and the road sections with the greatest number of possible overlapping routes. The following procedures are suggested:

i) Identify the origin-destination pairs of bicycle trips;

ii) Acquire at least four possible routes from each origin to the matching destination through the shortest path routines (using Dijkstra's algorithm);

iii) Identify roadway segments which concentrate a number of possible overlapping routes higher than one or more predetermined values by using selection tools;

iv) Design proposed cycling networks based on the roadway segments with the largest number of possible overlapping routes (item iii), on the volume and direction of motorized vehicle traffic flow (in this case, it is suggested avoiding roads with motorized vehicle traffic volume exceeding a predetermined limit, for example 120 vehicles/minute/lane) and roadway topography (it is suggested avoiding slopes greater than 5%, since they may undermine cycling trips).

The third and last step proposes a comparison between the obtained networks and the ones established by the municipal government, taking into account real and potential users. The inclusion of potential users is an additional step in relation to the exploratory work done by Guerrero et al. (2016). Initially, the method suggests using two assessment strategies, one using disaggregate trip data and the other using aggregate data to evaluate certain costs and benefits, both considering a band with a homogeneous width around the networks (for which a value of 400 meters is recommended, but this can be adjusted by the decision maker).

The first strategy (Evaluation I) was developed by accounting for the origin of the trips made by real and potential users included within the band around the networks. This was done using TransCAD´s "overlay" tool, which considered the overlapping of layers including the 400 meter band, the trip origins and the networks. This tool was also used to develop the second strategy (Evaluation II), considering the layers containing the 400 meter band, the census tracts and their aggregated population data, as well as the networks.

Afterwards, it is recommended to compare the networks considering cost criteria (C1 to C6) and benefit criteria (B1 to B6), described as follows:

C1. Total length of the network;

C2. Number of intersections per km over the network;

C3. Readability: Corresponds to the ease of assimilating the network by the user. It is characterized by the ratio between the number of conversions and the length of the route;

C4. Average distance of significant trip generators for cycling to the network -- in this case, it refers to the distance between schools (public and private) and the proposed networks;

C5. Average distance from the origin of trips to the network -- refers to the distance traveled by each user to access the proposed network;

C6. Percentage of trips made outside the network, represented by the percentage of the length of trips located outside the network to access the destination, relative to the total length of the network.

B1. Percentage of trip origins covered by the network;

B2. Percentage of trips made within the network, represented by the percentage of the length of trips covered within the network to access the destination, relative to the total length of the network;

B3. Percentage of the population served by the network (as informed by IBGE, i.e., the Census Bureau);

B4. Percentage of schools (public and private) served by the network;

B5. Percentage of commercial establishments;

B6. Percentage of industries.

As in the study of Milakis and Athanasopoulos (2014), we also consider the density of crossings (represented by cost C2), readability (cost C3) and centrality as evaluation criteria for cycling networks. Specifically regarding centrality, Milakis and Athanasopoulos (2014) considered trip generators within a 500 meter band around the studied networks. We used a 400 meter band as it corresponds to the maximum acceptable distance to access the proposed route (as suggested by

Ferraz and Torres, 2004, for a public transit case). Other criteria were used to indicate the effectiveness of the network or the capacity to serve a large number of users.

For the cost criteria, the higher the value, the worse the network service, either due to higher construction costs (C1), longer waiting times (C2), difficult user assimilation (C3) and greater distances travelled by the user (C4, C5 and C6). For the benefit criteria, the higher the value obtained, the better the service provided by the network. With the exception of B2, the percentages, in relation to the total of each item, were obtained by accounting for their items within the band around the proposed network.

A sensitivity analysis assuming minimum and maximum limits of cost and benefit criteria was conducted, in addition to the normalization of all obtained values to add to the criteria comparison.

4 APPLYING THE METHOD IN THE CASE STUDY

The disaggregated data used for this study was derived from a household origin-destination survey applied in the city of São Carlos in 2007-2008, considering a sample of 6% of the city´s population from that period, which corresponds to 19,784 trips (Rodrigues da Silva, 2008). The survey was developed in a joint effort of researchers from two local universities - the University of São Paulo and the Federal University of São Carlos - and the municipality. The data was collected from surveys conducted by trained staff, who asked the population sample about the trips they made during a typical weekday. From a total of 19,784 trips, 11,679 trips were considered appropriate for this study, after excluding cases in which important information was missing (for example, origin or destination variables).

 ACCEPTED MANUSCRIPT

Aggregated census data from 2010, when the city had a population of 212,465 inhabitants in the urbanized area (IBGE, 2015), was also used in the research. The data were available on a digital database maintained by the Brazilian census bureau (IBGE), which is responsible for collecting, treating and analysing economic and social data concerning all Brazilian municipalities and disseminating this information to the population. Most of the data provided by IBGE is geo-referenced. The data of origin-destination survey and the census data are available for public use.

4.1 Identifying potential users

The following variables were selected from the origin-destination survey dataset for the data-mining procedure: i) gender, ii) driver's license ownership, iii) mode choice, iv) socio-economic class, v) age, vi) distance between origin and destination, vii) if studying regularly, and viii) level of education. In this study, we used the SPSS Statistics program, developed by IBM, but other software can be used.

The variable associated to the transport mode was determined as dependent and binary in order to differentiate cycling from other modes. Therefore, the variable "bicycle mode" was assigned value 1 and the other modes value 0, making it easier to identify the transport mode used in the mining results. As for the independent variables, "gender" and "driver's license ownership" were also binary. For "distance", pre-calculated, non-binary values were chosen. The other independent variables ("socio-economic class", "if studying regularly", and "level of education") presented non-binary values obtained from the origin-destination survey classification, as presented in Table 1.

The Decision Tree technique associated with the CART algorithm (Classification and Regression Tree) was adopted for each sample (test and training). This technique presents a structure similar

 ACCEPTED MANUSCRIPT

to a tree, holding one root and a series of branches (intermediate and end). The tree root is called a "root node", which contains the original data. The intermediate branches are called "intermediate nodes." The end branches, from which no other branch is derived, are called the "end nodes" or "leafs." The intermediate nodes are derived from the root node by hierarchical decision rules (such as: "If… then…"). The Decision Tree technique has been used for studies in different fields with satisfactory results, as shown by Stojanova et al. (2006), Kheir et al. (2008), Xie et al. (2003), Rasouli and Timmermans (2014), Caballero et al. (2016), among others.

The CART algorithm made it easier to establish hierarchical decision rules for the original data using successive and binary divisions so as to make the resulting groups increasingly homogenous in relation to the dependent variable, as proposed by Breiman et al. (1984). The establishment of 50 'parent' nodes and 25 'children' nodes enabled us to identify the patterns of real cyclists and, from those, find out who the potential users are.

When data-mining was performed for the 11,679 trips, the trips already made by bicycle (mode = 1) and the ones made by other transport modes (mode = 0) could be identified. The trips presented a probability value, which corresponded to the probability of bicycle use by the correspondent users (potential users) and was conditional on the set of values of their independent variables. If this set of variables was close to the set of real users, i.e., came close to the pattern of real users, then the probability value associated with bicycle use would be higher for these users.

For both samples, the probabilities of bicycle use were exported as spreadsheet files and ranged from 0% to 30%. In order to obtain a representative number of potential users, those who presented a probability of using a bicycle value equal to or greater than 15% were considered

ACCEPTED MANUSCRIPT

potential users. Considering this refinement, in addition to excluding the real users (who already cycle), the sample obtained was equal to 258 trips (from more than the 11,000 initially considered trips), as presented in Figure 1.

4.2 Cycling network planning

As it was not possible to identify an origin-destination pair for the entire trip sample, at this stage, a total of 186 trips were considered. From these, four possible routes between each origin and destination were drawn, using the shortest path routines (with *TransCAD*, but other GIS could also be used) as shown in the example in Figure 2. It is reasonable to assume that one of these routes is similar to the actual route ridden by the respondent to his or her origin-destination pair.

After establishing the four possible routes between the origin-destination pairs, the roadway segments of the street network, which concentrated a number of possible overlapping routes higher than one or more predetermined values, were identified (item iii of the method). In this study, three values were considered: in the first case, all roadway segments with over 5 overlapping routes; in the second case, those with more than 10 overlapping routes; and in the third case, those with more than 15 overlapping routes. These networks are referred to as Network_{5+}, Network_{10+} and Network_{15+}, respectively (Figure 3 a). In case there were two or more roadways with the same number of possible overlapping routes higher than the pre-established values, the choice was made for the one with lowest motorized-vehicle traffic volume. Moreover, whenever possible, we avoided establishing networks in the opposite direction to the existing motorized-vehicle flow. With regard to topography, in many situations the analysis of the longitudinal profile to define the route of the cycling network was carried out using *Google*

Earth, as suggested by Guerreiro and Rodrigues da Silva (2013) and Guerreiro et al. (2013). In addition to the proposed cycling networks, those designed by the municipal government of São Carlos were also considered: the existing cycling network (ECN) and the designed cycle network (DCN) (Figure 3 b).

4.3 Comparative analysis of the cycling networks

Concerning the two assessment strategies (Evaluation I and Evaluation II), 186 trips were accounted to real users and 258 to potential users, as shown in Table 2. The total urban population considered, according to 2010 Census data, was equal to 212,465 (IBGE, 2015) and the total length of the city's street network was equal to 1,088.66 km.

Considering Evaluation I (disaggregated data), the existing network (ECN) was, compared to the networks proposed here, the furthest from the origin of the users' trips: for Network 5+, 7.0 times more distant in the case of real users and 4.5 times for potential users. Regarding the network designed by the municipality (DCN), this ratio was around 2.0 for both types of users. It should be noted that the farther away the cycling network is from the origin of the user's trips, real or potential, the higher the risk of traffic accidents, given that the users are exposed to unsafe traffic conditions until they access the networks.

The ECN also had the lowest percentage of trips made within the network. The coverage percentage of the DCN for a 400 meter band was close to that obtained for Network_{15+}, especially in the case of potential users (55% vs. 60%), even though the designed network is 31% shorter than the proposed network.

Using the aggregated population data in census tracts (Evaluation II), both for real users and potential ones, Network_{5+} presented the highest percentage of population coverage (78%)

compared to other networks due to its greater extension and comprehensiveness. On the other hand, the ECN showed the lowest percentage of the covered population (16%), followed by DCN (41%), as shown in Table 2.

Table 3 shows the comparison of the studied networks, according to the cost and benefit criteria, considering real and potential users.

The only criteria that varied in the comparison between real and potential users were the cost criteria C5 and C6 and the benefit criteria B1 and B2. Regarding the cost criteria, real users had higher C5 values for ECN and DCN networks and higher C6 values for Network_{10+}. At the same time, real users had higher B1 and B2 values for nearly all the proposed networks, while potential users had higher percentages for ECN and DCN networks.

Thus, for the disaggregated data, for both real and potential users, Network_{5+} resulted in the lowest cost values of C4, C5 and C6, while ECN resulted in the highest values of these cost variables. Still considering this network, all the benefits were less than those presented by other networks, notably Network_{5+}, which is the network with higher extension and coverage.

As for the data regarding the population as a whole (aggregated), while DCN, designed by the municipality, was about 31% shorter in length than Network_{15+} (C1 in Table 3), it attended only 16% fewer people (B3 in Table 3). The sensitivity analysis was performed considering the limits given in Table 4.

The normalization was performed considering the 0 to 1 interval. For the costs, the lower the obtained value, the better the alternative, and, therefore, value 1 was assigned. For the highest value in the analysis, 0 was assigned. In contrast, for the benefits, the higher the percentage and values, the better the alternative, i.e., value 1 was assigned for the highest percentage and 0 was

 ACCEPTED MANUSCRIPT

assigned for the lowest percentage. Considering cost C1, for example, the limits assumed in the normalization were 10 (minimum) and 300 (maximum), which is, the normalization process assigned 1 to the minimum limit and 0 to the maximum limit. The values of C1 for the considered networks were situated between the minimum and maximum limits, varying from 10.53 to 95.01 km (networks DNC and $Network_{5+,}$ respectively). In other words, it was necessary to interpolate the values of C1 from the studied networks as they presented intermediate values between 0 and 1. Therefore, 0.71 was obtained as a value for $Network_{5+}$, for example. The same procedure was performed for the other costs and all of the benefits. The results are shown in Figure 4.

By analysing cost variable C1, it could be observed that for both real and potential users, there is a close relationship between the normalized values for the networks DCN, $Network_{5+}$, $Network_{10+}$ and $Network_{15+}$. On the other hand, it is clear that ECN presents the best condition as it has the lowest cost value C1.

The normalized values obtained for the benefits of ECN were lower than the values obtained in any of the other studied networks, with the exception of the cost variables C1, C2 and C3. This happens because ECN has less coverage when compared to other networks (in the case of benefits), as well as lower construction costs (C1), lower travel time (C2) and easier network assimilation (C3), all of which can be considered good for the user.

5 CONCLUSIONS

This study presented the development and application of an objective method to design and compare cycling networks, considering data from an origin-destination survey (i.e., disaggregated data). The proposed approach was presented as an alternative for methods that use aggregated data (such as in census tracts or traffic zones) which implicitly assume a homogeneous distribution of the population in the considered areas. In addition, a procedure to identify potential bicycle users and a comparison method considering cost and benefit criteria, common to all studied networks, was also discussed. Especially regarding the cycling networks, in addition to the three alternatives proposed (Network$_{5+}$, Network$_{10+}$ and Network$_{15+}$), the existing cycling network (ECN) and the designed cycling network (DCN - designed by the municipal government of São Carlos) were also analysed.

Judging by the results of the case study, the method used to identify potential users of cycling transport can be considered a valid alternative, in particular when disaggregated data of an origin-destination survey are available. Thus, in order to use the method presented here, it is not necessary to adopt stated or revealed preference surveys, answering the first cycling network planning research question presented at the beginning of the study. Furthermore, calculating the probabilities of potential users riding bicycles enabled us to identify the users who were likely to use this transportation mode. This identification, associated with disaggregated information of the users, results in further refinement of the obtained outcomes.

Although it is not as dynamic as the method proposed by Lovelace et al. (2015), using data from an origin-destination survey seems to be effective to plan and design cycling networks as accurate location data from real users may help to identify suitable routes between origins and destinations, which would not happen if an exploratory procedure was chosen. Using the exact geographical location enabled us to include the spatial analysis of the users, contributing towards a more realistic delineation of the networks, as close as possible to the origins and destinations.

ACCEPTED MANUSCRIPT

The effectiveness of the method demonstrated that it is possible to plan for cycling networks using disaggregated population data, positively responding to the second research question raised at the beginning of the research. In addition, subject to the availability of geo-referenced household origin-destination survey data and information on users' transportation modes, the method proposed in this study can be used in any city that aims to establish or expand a cycling network. The method can also be applied to larger trip generators, for instance, university campuses or shopping malls.

In a comparison of the networks considering the proposed evaluation strategies, the exact location of potential users and their destinations helped us to obtain the values of the same variables obtained for real users. A direct comparison between the two groups could be made, which would not be possible if they were only aggregate population data. In addition, using disaggregated data enabled us to measure and analyse the proximity of users to the networks, as well as the number of people served by them, while using aggregate data it would only be possible to identify the number of people served, with no information of their location and type of users (real or potential).

If the values obtained for the cost/benefit criteria considering the actual and potential users were very different from each other, this could be an indicative of a great difference in the spatial distribution between them, which would consequently lead to poor service of the networks for one of two categories of users. Another possibility is related to the choice of variables used in data mining, suggesting that it may be necessary to choose other variables considered more important for choosing the bicycle as a means of transport.

 ACCEPTED MANUSCRIPT

Regarding the costs and benefits considered in the evaluation of the proposed cycling networks, the selected criteria were not only related to the physical characteristics of the networks (such as extension, number of turns per km, number of turns per network extension), but also criteria related to the exact location of real and potential users, as well as the location of some trip generators. The selected criteria were all effective (as also observed in the study by Milakis and Athanasopoulos, 2014) as they considered a direct comparison between the proposed networks and the existing infrastructure.

In addition, the sensitivity analysis associated with the normalization of the cost and benefit values obtained enabled us to understand the relationship better between all the studied networks, including the proposed networks. Moreover, the quality of the ECN for real and potential users can clearly be observed.

Even though not all of the cost and benefit values were originally set to the same scale, the normalization procedure based on the same minimum and maximum limits allowed for a comparison using the 0-1 scale, which enabled a better understanding of the relationship among the proposed networks and the ones designed by the municipality.

Not always is a wide network totally beneficial for bicycle users, even though it has many benefits. One reason for this may be linked to the characteristics of the road network (number of intersections) related to the length and readability of the network (network assimilation is user friendly). This set of features can increase users’ travel time and requires a larger and more detailed project of horizontal and vertical traffic signs and traffic control devices with high implementation and maintenance costs.

 ACCEPTED MANUSCRIPT

In general, as in the city of São Paulo (BBC, 2015), the city of São Carlos currently has many discontinuous sections of the cycle network, which directly reflects the results obtained from the ECN. Furthermore, as shown in a study conducted by Krizek and Roland (2005), one of the discontinued paths of ECN is also located on the left side of the roadway, also ending in an intersection. This configuration causes great discomfort for motorized vehicle drivers, but mainly for the cyclists. Due to the ECN path ending, cyclists need to cross two flows of existing traffic before they can position themselves on the right side of the roadway (in the direction of the traffic flow) and continue their journey. In addition to the discomfort experienced by cyclists, this situation causes great insecurity and may be one of the reasons that contribute to the poor service of the network in question. A cycling network that does not provide a direct connection between origin and destination points for riders cycling to work can also reduce the use of the infrastructure. This is, however, particularly the case of riders cycling to work, and not of all cyclists, as observed by Standen et al. (2017).Considering this scenario, it can be affirmed that the current stage of the existing network provides little or no safety for cyclists and does not reduce significant travel time for bicycle trips. This situation can be explained by the lack of objective methods for network design and the disregard of potentially relevant data, resulting in a cycling infrastructure that possibly serves political and electoral purposes. By implementing the designed network (DCN), there may be an improvement in the service for users, especially for potential ones as the results were very similar to those obtained in one of the proposed networks. However, the results from the networks designed by the municipality are far from those obtained from the networks proposed by this study, which would ensure a better service. Also, the implementation of any of the proposed networks would eventually stimulate the development of a bicycle sharing system, as observed in the study conducted by Faghih-Imani e Eluru (2016). Still another approach that can add to the analysis of costs and benefits is the Network Robustness Index (NRI) developed by Burke e Scott (2016), that can be used to identify the bicycle facility design that limits traffic disruption for any road link in an urban network.

Throughout this study, a comprehensive overview of the current situation of the cycling transportation system in the city of São Carlos was provided. However, it should be noted that the variables considered do not exhaust the possibilities for evaluating and comparing proposed and existing cycling networks. While this may indicate a weakness in the method, it certainly does not invalidate it. From this standpoint, adopting new cost and benefit variables related to network characteristics may contribute to a better understanding of the relationship between them. Another important detail that is still open for investigation is the development of strategies that can help decision-makers to formally determine which one is the best alternative to put into practice when comparing proposed networks that generally perform well (as the three alternatives proposed in this study).

Acknowledgments

This research was supported by CNPq (Proc. 141290/2011-0) and by FAPESP (Proc. 2015/50129-5).

6 REFERENCES

Aultman-Hall, L.; Hall, F.L.; Baetz, B.B. (1997) Analysis of bicycle commuter routes using geographic information systems implications for bicycle planning. *Transportation Research Record*, n. 1578, pp. 102-110.

BBC - British Broadcasting Corporation (2015) Pedal power forges ahead in Sao Paulo's famous Avenida Paulista. Latin America & Caribbean. Available from http://www.bbc.com/news/world-latin-america-33271772. Accessed July 17, 2015.

Benicchio T. (2012) Bicicletas transformando a cidade [Bicycles transforming the city]. In: Padovano, B.R.; Sala, P.B.; Namur, M. (Org.) São Paulo: em busca da sustentabilidade. São Paulo: PINI - EDUSP, v. 1, p. 216-230.

Bergman, C.; Oksanen, J. (2016) Conflation of OpenStreetMap and mobile sports tracking data for automatic bicycle routing. *Transactions in GIS*, v. 20, n.6, p. 848-868. DOI: 10.1111/tgis.12192.

Bergström, A.; Magnusson, R. (2003) Potential of transferring car trips to bicycle during winter. *Transportation Research Part A*, v. 37: p. 649-666. DOI: 10.1016/S0965-8564(03)00012-0.

BHTRANS - Empresa de Transportes e Trânsito de Belo Horizonte. 2012. Plano de mobilidade urbana de Belo Horizonte [Urban mobility plan of Belo Horizonte]. Belo Horizonte, Brazil.

Boettge, B.; Hall, D. M.; Crawford, T. (2017) Assessing the bicycle network in St. Louis: A place-based user-centered approach. *Sustainability*, v. 9, n. 2, p. 241. DOI:10.3390/su9020241.

Boussauw, K.; van Meeteren, M.; Witlox, F. (2014) Short trips and central places: The home-school distances in the Flemish primary education system (Belgium). *Applied Geography*, n. 53, p. 311-322. DOI: 10.1016/j.apgeog.2014.06.025

Burke, C. M.; Scott, D. M. (2016) The space race: A framework to evaluate the potential travel-time impacts of reallocating road space to bicycle facilities. *Journal of Transport Geography*, n. 56, p. 110-119. DOI: http://dx.doi.org/10.1016/j.jtrangeo.2016.09.004.

Brasil (2001) Estatuto da Cidade [City´s statute]. Law number 10.257 (10 jul. 2001).

Brasil (2013) Projeto de Lei n° 4997/2013. Dispõe sobre o IPI - Imposto sobre Produtos Industrializados, fixando alíquota de zero por cento para as bicicletas (item 8712.00.10 da NCM - Nomenclatura Comum do Mercosul), suas partes e acessórios [About tax exemption for the production of bicycles (item 8712.00.10 of the NCM - Mercosur Common Nomenclature), parts and accessories]. Available from http://www.camara.gov.br/proposicoesWeb/fichadetramitacao?idProposicao = 565266 (accessed July 16, 2015).

Brasil (2009) Projeto de Lei n° 6474/2009. Institui o Programa Bicicleta Brasil, para incentivar o uso da bicicleta visando a melhoria das condições de mobilidade urbana [Establishing the Bicycle Program Brazil, to encourage the use of bicycles aiming at improving the conditions of urban mobility]. Available from http://www.camara.gov.br/proposicoesWeb/prop_mostrarintegra?codteor = 717079&filename = PL+6474/2009 (accessed July 16, 2015).

Brasil (2013) Projeto de Lei n° 6418/2013. Altera a Lei n° 7.418, de 16 de dezembro de 1985, para permitir que o vale-transporte seja utilizado para o custeio de transporte do empregado por meio de bicicletas [modifies Law No. 7,418, of December 16, 1985, to allow employees to use the travel-voucher also for bicycle trips]. Available from http://www.camara.gov.br/proposicoesWeb/prop_mostrarintegra?codteor = 1142479&filename = PL+6418/2013 (accessed July 16, 2015).

Brasil (2013) Projeto de Lei n° 6724/2013. Acrescenta artigo à Lei n° 7.418, de 16 de dezembro de 1985, para instituir o "Vale-Bicicleta" [modifies Law No. 7,418, of December 16, 1985, to establish a travel-voucher for bicycles]. Available from http://www.camara.gov.br/proposicoesWeb/prop_mostrarintegra?codteor = 1176752&filename = PL+6724/2013 (accessed July 16, 2015).

Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. (1984) Classification and Regression Trees. Wadsworth International Group, Califórnia.

Caballero, D.; Caro, A.; Rodríguez, P. G.; Duran, M. L.; Avila, M. M.; Palacios, R.; Antequera, T.; Perez-Palacios, T. (2016) Modeling salt diffusion in Iberian ham by applying MRI and data mining. *Journal of Food Engineering*, v. 189, p. 115-122. DOI: 10.1016/j.jfoodeng.2016.06.003.

Câmara dos Deputados (2014). Comissões vão avaliar regras tributárias para bicicletas [Commissions will assess tax rules for bicycles]. Câmara Notícias. Available from http://www2.camara.leg.br/camaranoticias/noticias/TRANSPORTE-E-TRANSITO/467797-COMISSOES-VAO-AVALIAR-REGRAS-TRIBUTARIAS-PARA-BICICLETAS.html (accessed July 16, 2015).

Cardoso, C. P. (2015) A contribuição da malha viária no âmbito da cidade de Aracaju [The road network contribution in the city of Aracaju]. *Revista Ideias & Inovação*, v. 2, p. 23-34. Aracaju.

Castillo-Manzano, J. I.; López-Valpuesta, L.; Sánchez-Braza, A. (2016) Going a long way? On your bike! Comparing the distances for which public bicycle sharing system and private bicycles are used. *Applied Geography*, v. 71, p. 95-105. DOI: 10.1016/j.apgeog.2016.04.003.

DENATRAN - Departamento Nacional de Trânsito (2015) Vehicle fleet numbers. Available from http://www.denatran.gov.br/index.php/estatistica/237-frota-veiculos (accessed July 17, 2015).

Dill, J.; Voros, K. (2007) Factors affecting bicycling demand initial survey findings from the Portland, Oregon, Region. *Transportation Research Record,* n. 2031, p. 9-17. DOI: 10.3141/2031-02.

Faghih-Imani, A.; Eluru, N. (2016) Determining the role of bicycle sharing system infrastructure installation decision on usage: Case study of Montreal BIXI System. *Transportation Research Part A*, n. 94, p. 685-698. DOI:10.1016/j.tra.2016.10.024.

Ferrari, R. (2008) Plano diretor cicloviário integrado de Porto Alegre - a questão da institucionalização [Integrated master cycle plan of Porto Alegre - the issue of institutionalization]. XXIII ANPET - Congresso de Pesquisa e Ensino em Transportes, November 2008, Fortaleza, Brasil.

Ferraz, A.C.P; Torres, I.G.E. (2001) Transporte Público Urbano. São Carlos: Rima.

Guerreiro, T.C.M.; Rodrigues da Silva, A.N. (2013) A computer-aided approach for planning sustainable trips to large trip generators: the case of cycling routes serving university Campuses. In: Geertman S.; Toppen, F. and Stillwell, J. (org.). Lecture Notes in Geoinformation and Cartography, Springer Berlin Heidelberg, 1st ed. Pp. 297-308.

Guerreiro T.C.M.; Stein P.P.; Rodrigues da Silva, A.N. (2013) Potencial de uma infraestrutura cicloviária para diferentes usuários de um polo gerador de viagens: o caso de um campus universitário [Potential of cycle routes for different users of trip generators: the case of an university campus]. XXVII ANPET - Congresso de Pesquisa e Ensino em Transportes, November 2013, Belém, Brasil.

Guerreiro, T.C.M.; Rodrigues da Silva, A.N.; Ramos, R.A.R. (forthcoming) Explorando dados de pesquisas origem-destino para o planejamento de novas redes cicloviárias[Exploring data origin-destination survey for the planning new cycling networks]. Revista dos Transportes Públicos.

Heinen, E.; van Wee, B.; Maat, K. (2010) Commuting by bicycle: An overview of the literature. *Transport Reviews: A Transnational Transdisciplinary Journal*, v. 30, p. 59-96. DOI: 10.1080/01441640903187001.

Huang, Y; Ye, G. (1995) Selecting bicycle commuting routes using GIS. *Berkeley Planning Journal*, v. 10, p. 75-90.

Hsu, T.P.; Lin, Y.T. (2012) A model for planning a bicycle network with multicriteria suitability evaluation using GIS. In: Brebbia, C.A.; Zubir, S.S. (ed.). Management of Natural Resources, Sustainable Development and Ecological Hazards III. Southampton: WIT Press, Section 4: Planning and development, p. 243-252.

IBGE - Instituto Brasileiro de Geografia e Estatística. 2015. Projeção da população das unidades da federação por sexo e idade: 2000-2030 [Projection of the population of the federation units by gender and age: 2000-2030]. Available from http://www.ibge.gov.br/home/estatistica/populacao/projecao_da_populacao/2013/default_tab.shtm (accessed July 17, 2015).

IDU - Instituto de Desarrollo Urbano (2014) Espacio público [Public space]. Available from http://www.idu.gov.co/web/guest/espacio_ciclo_plan (accessed July 13, 2015).

ACCEPTED MANUSCRIPT

Implan - Instituto Municipal de Planeación Urbana San Pedro Garza García (2012) Plan de desarrollo urbano municipal de San Pedro Garza García, N.L. 2030 - Versión para consulta pública [Municipal urban development plan for San Pedro Garza Garcia, N.L. 2030 - Version for public consultation]. Available from http://www.implansp.org/docs/LIBRO.pdf (accessed July 13, 2015).

IPPUC - Instituto de Pesquisa e Planejamento Urbano de Curitiba. 2013. Plano estratégico cicloviário de Curitiba [Strategic cycling plan of Curitiba].

Kheir, R.B.; Chorowicz, J.; Abdallah, C.; Dhont, D. (2008) Soil and bedrock distribution estimated from gully form and frequency: A GIS-based decision-tree model for Lebanon. *Geomorphology*, n. 93, p. 482-492. DOI: 10.1016/j.geomorph.2007.03.010.

Kirner J. (2006) A method for the definition of cycling routes in urban areas (in Portuguese). Master's thesis, Federal University of São Carlos, São Carlos, Brazil.

Krizek, K.J.; Roland, R.W. (2005) What is at the end of the road? Understanding discontinuities of on-street bicycle lanes in urban settings. *Transportation Research D*, v. 10, p. 55-68. DOI: 10.1016/j.trd.2004.09.005.

Larsen, J.: Patterson, Z.; El-Geneidy, A. (2013) Build it. But where? The use of Geographic Information Systems in identifying locations for new cycling infrastructure. *International Journal of Sustainable Transportation*, v. 7, p. 299-317. DOI: 10.1080/15568318.2011.631098.

Lovelace, R.; Goodman, A.; Aldred, R.; Berkoff, N.; Abbas, A.; Woodcock, J. (2015) The Propensity to Cycle Tool: An open source online system for sustainable transport planning. Available from http://arxiv.org/abs/1509.04425 (accessed January 13, 2015).

Mackett, R.L. (2001) Policies to attract drivers out of their cars for short trips. *Transportation Research Policy*, v. 8, p. 295-306.

Medeiros, R.M.; 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*, v. 1, p. 28-44. Doi:10.1080/21650020.2013.866875.

Milakis, D.; Athanasopoulos, K. (2014) What about people in cycle network planning? Applying participative multicriteria GIS analysis in the case of the Athens metropolitan cycle network. *Journal of Transport Geography*, v. 35, p. 120-129.

Neri, T.B. (2012) Cycling system potentials: A case study of Maringá (in Portuguese). Master's thesis, State University of Maringá. Maringá, Brazil.

Ortúzar, J.D.; Iacobelli, A.; Valeze, C. (2000) Estimating demand for a cycle-way network. *Transportation Research Part A*, v. 34, p. 353-373.

Plataforma Urbana (2015) Los planes de ciclovías para el Gran Santiago hasta 2032 [Cycleway plans for Greater Santiago until 2032]. Available from http://www.plataformaurbana.cl/archive/2015/02/03/los-planes-de-ciclovias-para-el-gran-santiago-hasta-2032/ (accessed July 13, 2015).

PDTU (2010) Plano Diretor de Transporte Urbano e Mobilidade do Distrito Federal e Entorno - PDTU/DF [Master Plan for Urban Transport and Mobility in the Federal District and surrounding areas - PDTU /DF].

Prefeitura do Município de São Paulo (2015) Plano de mobilidade de São Paulo - Modo bicicleta [Mobility plan of São Paulo - Cycling mode] Available from http://www.prefeitura.sp.gov.br/cidade/secretarias/upload/chamadas/planmobsp-rev002_1428005731.pdf (accessed July 14, 2015).

Rasouli, S.; Timmermans, H. (2014) Effects of travel time delay on multi-faceted activity scheduling under space-time constraints: A simulation study. *Travel Behaviour and Society*, v. 1, p. 31-35. DOI: 10.1016/j.tbs.2013.10.002.

Rodrigues da Silva, A.N. (2008) Origin-destination survey of São Carlos. Report of the research project FAPESP 2004/15843-4. São Carlos, Brazil.

Rybarczyk, G., Wu, C. (2010) Bicycle facility planning using GIS and multicriteria decision analysis. *Applied Geography*, v. 30, p. 282-293.

Saghapour, T.; Moridpour, S.; Thompson, R. (2017) Measuring cycling accessibility in metropolitan areas. *International Journal of Sustainable Transportation*, v. 11, n. 5, p. 381-394. DOI: 10.1080/15568318.2016.1262927.

Segadilha, A.B.P. (2014) Identification of factors that influence the choice of route for cyclists: a case study of São Carlos (in Portuguese). Master's thesis, Federal University of São Carlos. São Carlos, Brazil.

SMAC - Secretaria Municipal de Meio Ambiente (2010) Programas cicloviários - Rio, capital da bicicleta. Available from http://www.rio.rj.gov.br/web/smac/rio-capital-da-bicileta (accessed July 14, 2015).

Sousa, I.C.N.; Sanches, S.P. (2015) Proposta de procedimentos para definição de redes cicloviárias em cidades de porte médio [Proposed procedures for defining cycling networks in medium-sized cities]. XXIX ANPET - Congresso de Pesquisa e Ensino em Transportes, Novembro 2015, Ouro Preto, Brasil.

Standen, C.; Crane, M.; Collins, A.; Greaves, S.; Rissel, C. (2016) Determinants of mode and route change following the opening of a new cycleway in Sydney, Australia. *Journal of Transport & Health*, v. 4, p. 255-266. DOI: 10.1016/j.jth.2016.10.004.

Stojanova, D.; Panov, P.; Kobler, A.; Džeroski, S.; Taškova, K. (2006) Learning to predict forest fires with different data mining techniques. In: Proceedings of the 9th International multi conference Information Society IS 2006, October 2006. Jožef Stefan Institute, Ljubljana.

Vá de Bike (2014) Reduzir impostos da bicicleta aumentaria a arrecadação (além de deixá-la mais barata) [Reductions in bicycle taxes would increase revenue (besides making bycicles cheaper]. Available from http://vadebike.org/2013/11/bicicleta-para-todos-reducao-impostos/ (accessed July 16, 2015).

Wang, J.Y.T.; Mirza, L., Cheung, A.K.L.; Moradi, S. (2014) Understanding factors influencing choices of cyclists and potential cyclists: A case study at the University of Auckland. *Road & Transport Research*, v. 23, p. 37-51.

Wooliscroft, B.; Ganglmair-Wooliscroft, A. (2014) Improving conditions for potential New Zealand cyclists: An application of conjoint analysis. *Transportation Research Part A*, v. 69, p. 11-19. DOI: 10.1016/j.tra.2014.08.005.

Wuerzer, T.; Mason, S. G. (2015) Cycling willingness: Investigating distance as a dependent variable in cycling behavior among college students. *Applied Geography*, v. 60, p. 95-106. DOI: 10.1016/j.apgeog.2015.03.009.

Xie, C.; Jinyang, L.; Parkany, E. (2003) Work travel mode choice modeling with data mining: decision trees and neural networks. *Transportation Research Record: Journal of the Transportation Research Board*, v. 1854, p. 50-61. DOI: 10.3141/1854-06.

Yamashita, Y.; Dantas, A.; Taco, P.; Yamamoto, K. (1998) Geographical Information System & Remote Sensing in bicycle planning. Proceedings of 21th Annual Infrastructure Planning Conference, pp. 295-298, Shiga, Japan.

Zhang, D.; Magalhães, D.J.A.V.; Wang, X.C. (2014) Prioritizing bicycle paths in Belo Horizonte City, Brazil: Analysis based on user preferences and willingness considering individual heterogeneity. *Transportation Research Part A*, v. 67, p. 268-278. DOI: 10.1016/j.tra.2014.07.010.

Table 1: Description of non-binary variables extracted from the O/D survey

Variable	Possible values
Social-economic class	A1. A2.B1. B2. C1. C2. D and E. where
MW * = R$ 380.00 (around US$162.00)	A1: maximum income < 26 MW
	A2: maximum income < 17.5 MW
	B1: maximum income < 9.5 MW
	B2: maximum income < 5.5 MW
	C1: maximum income < 3.5 MW
	C2: maximum income < 2.0 MW
	D: maximum income < 1.5 MW
	E: maximum income < 1 MW
If studying regularly	1. 2. 3. 4. 5. 6 and 7. where:
	(1) No
	(2) Early childhood - Preschool
	(3) Primary school - 1st to 4th Grade
	(4) Middle school - 5th to 8th Grade
	(5) High School
	(6) University
	(7) Others
Level of education	1. 2. 3. 4. 5. 6. 7 and 8. where:
	(1) No formal education
	(2) Preschool
	(3) Incomplete middle school
	(4) Complete middle school
	(5) Incomplete high school
	(6) Complete high school
	(7) Incomplete university
	(8) Complete university

(*) The minimum wage (MW) in Brazil, which is set every year by the federal government, is the lowest monthly salary an employer can legally pay to their employees for a full time job.

 ACCEPTED MANUSCRIPT

Table 2: Variables calculated to evaluate the cycling networks - Real users/*potential users*

EVALUATION I					
Based on data obtained from origin-destination survey					
Evaluation data	**Networks**				
	Network$_{5+}$	**Network $_{10+}$**	**Network $_{15+}$**	**ECN**	**DCN**
Total number of origins	140/*275*	140/*275*	140/*275*	140/*275*	140/*275*
Average distance from origin to network (m)	167.38/ *236.92*	208.51/ *279.31*	309.10/ *401.12*	1203.67/ *1017.87*	685.94/ *492.38*
Trip origins contained in a 400 m band	122 (87%)/ *234 (85%)*	110 (79%)/ *215 (78%)*	90 (64%)/ *166 (60%)*	22 (16%)/ *59 (21%)*	64 (46%)/ *151 (55%)*
Total length of trips (km)	734.70/ *1014.67*	760.57/ *1067.03*	815.68/ *1079.32*	631.61/ *831.67*	644.69/ *878.41*
Total trip length inside the network (km)	680.14 (93%)/ *933.12 (92%)*	692.27 (91%)/ *975.13 (91%)*	721.31 88%)/ *951.13 (88%)*	60.37 (10%)/ *148.48 (18%)*	244.72 (38%)/ *443.70 (51%)*
Total trip length outside the network (km)	54.56 (7%)/ *81.55 (8%)*	68.30 (9%)/ *91.91 (9%)*	94.37 (12%)/ *128.19 (12%)*	571.24 (90%)/ *683.19 (82%)*	399.97 (62%)/ *434.71 (49%)*
Total network extension (km)	95.01	74.48	60.94	10.53	42.32
% total street network	8.73	6.84	5.60	0.97	3.89
	EVALUATION II				
	Based on population data				
Evaluation data	**Networks**				
	Network$_{5+}$	**Network $_{10+}$**	**Network $_{15+}$**	**ECN**	**DCN**
Population (inhab.) (Census /IBGE) contained in a 400 m band	166.272	150.974	122.407	33.446	88.093
% total urban population	78	71	58	16	41

Table 3: Direct comparison of cost and benefit criteria - Real users/*potential users*

Networks / Variables	ECN	DCN	Network$_{15+}$	Network$_{10+}$	Network$_{5+}$
COSTS					
Total length of the network (km) (C1)	10.53	42.32	60.94	74.48	95.01
Number of intersections per km over the network (C2)	7.69	13.52	18.97	15.90	15.65
Ratio between the number of conversions and the length of the route (C3)	1.14	1.56	6.86	3.88	3.48
Average distance between schools (public and private) and the proposed networks (C4)	1001.52	575.12	310.19	225.98	185.84
Average distance from the origin of trips to the network (C5)	1203.67/ *1017.87*	685.94/ *492.38*	309.10/ *401.12*	208.51/ *279.31*	167.38/ *236.92*
Percentage of the length of trips located outside the network to access the destination (C6)	90.44%/ *82.15%*	62.04%/ *49.49%*	11.57%/ *11.88%*	8.98%/ *8.61%*	7.43%/ *8.04%*
BENEFITS					
Percentage of trip origins covered by the network (B1)	15.71%/ *21.45%*	45.71%/ *54.91%*	64.29%/ *60.36%*	78.57%/ *78.18%*	87.14%/ *85.09%*
Percentage of trips made within the network (B2)	9.56%/ *17.85%*	37.96%/ *50.51%*	88.43%/ *88.12%*	91.02%/ *91.39%*	92.57%/ *91.96%*
Percentage of the population served by the network (IBGE) (B3)	15.74%	41.46%	57.61%	71.06%	78.26%
Percentage of schools (public and private) served by the network (B4)	16.16%	42.42%	68.69%	78.79%	83.84%
Percentage of commercial establishments (B5)	17.83%	48.92%	88.18%	91.93%	94.03%
Percentage of industries (B6)	21.40%	39.74%	79.04%	84.28%	88.65%

 ACCEPTED MANUSCRIPT

Table 4: Limits adopted in the normalization process of the results

Limits	C1	C2	C3	C4	C5	C6
Minimum	10	5	0	100	100	0%
Maximum	300	50	20	1500	1500	100%
	B1	B2	B3	B4	B5	B6
Minimum	0%	0%	0%	0%	0	0
Maximum	95%	95%	95%	95%	95%	95%

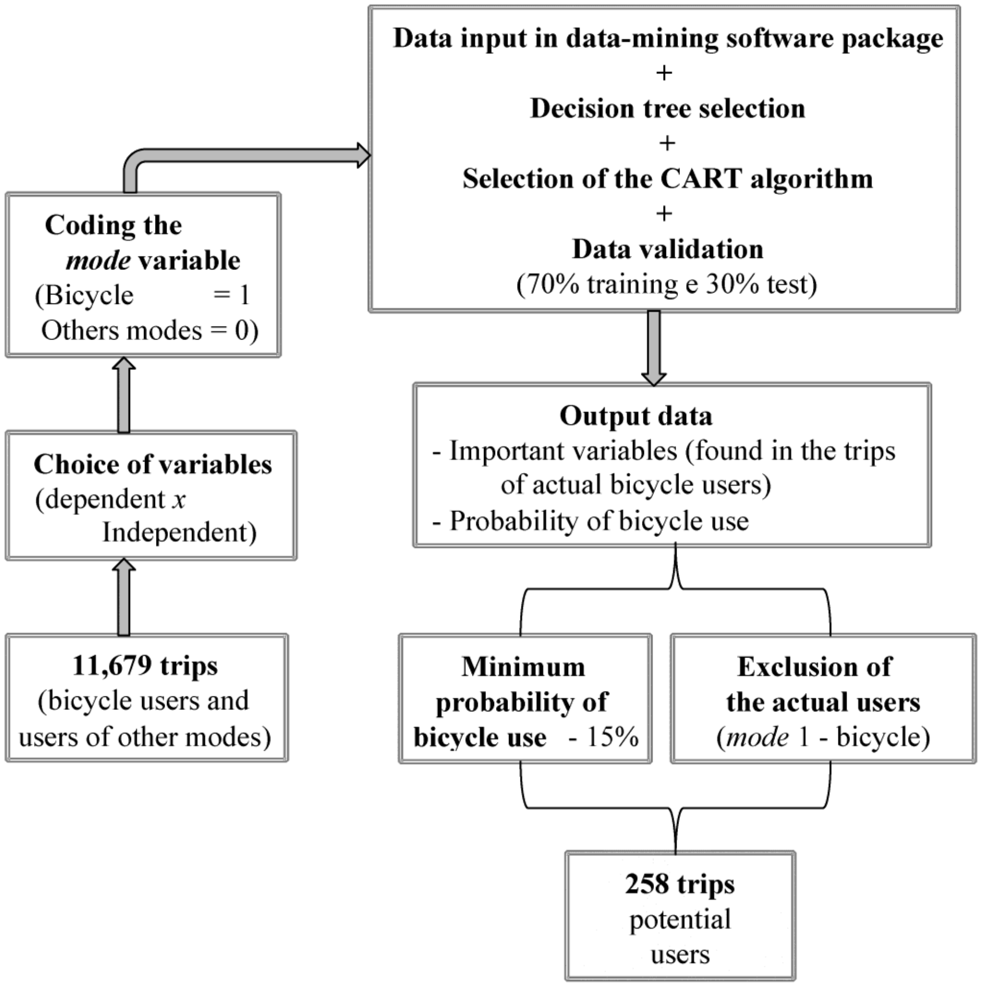

Figure 1: Summary of the procedure used to select potential users of the bicycle mode

ACCEPTED MANUSCRIPT

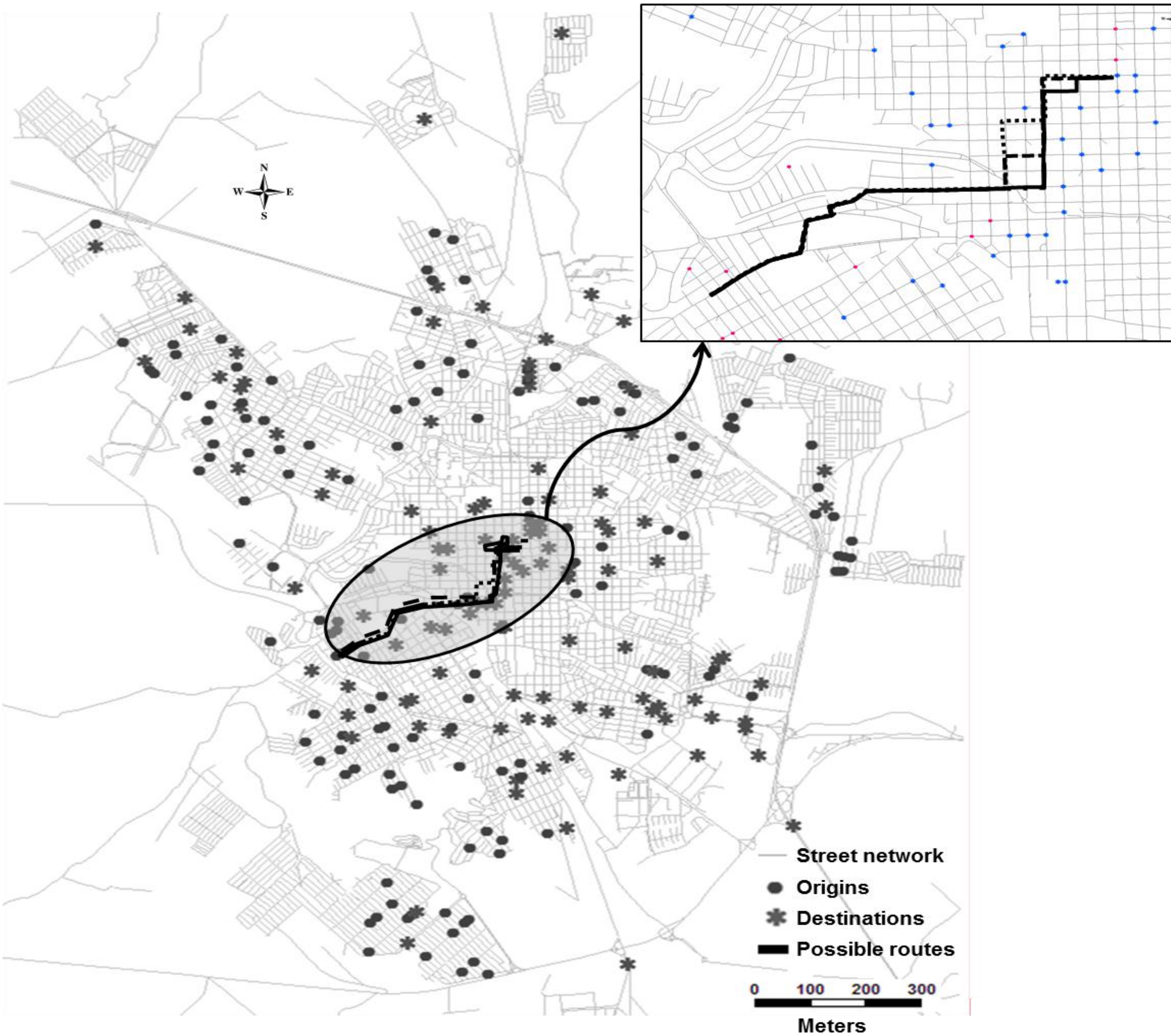

Figure 2: Georreferencing of bicycle trip origins and destinations and an example of four possible routes between one origin-destination pair. Source: Guerreiro et al. (2016)

 ACCEPTED MANUSCRIPT

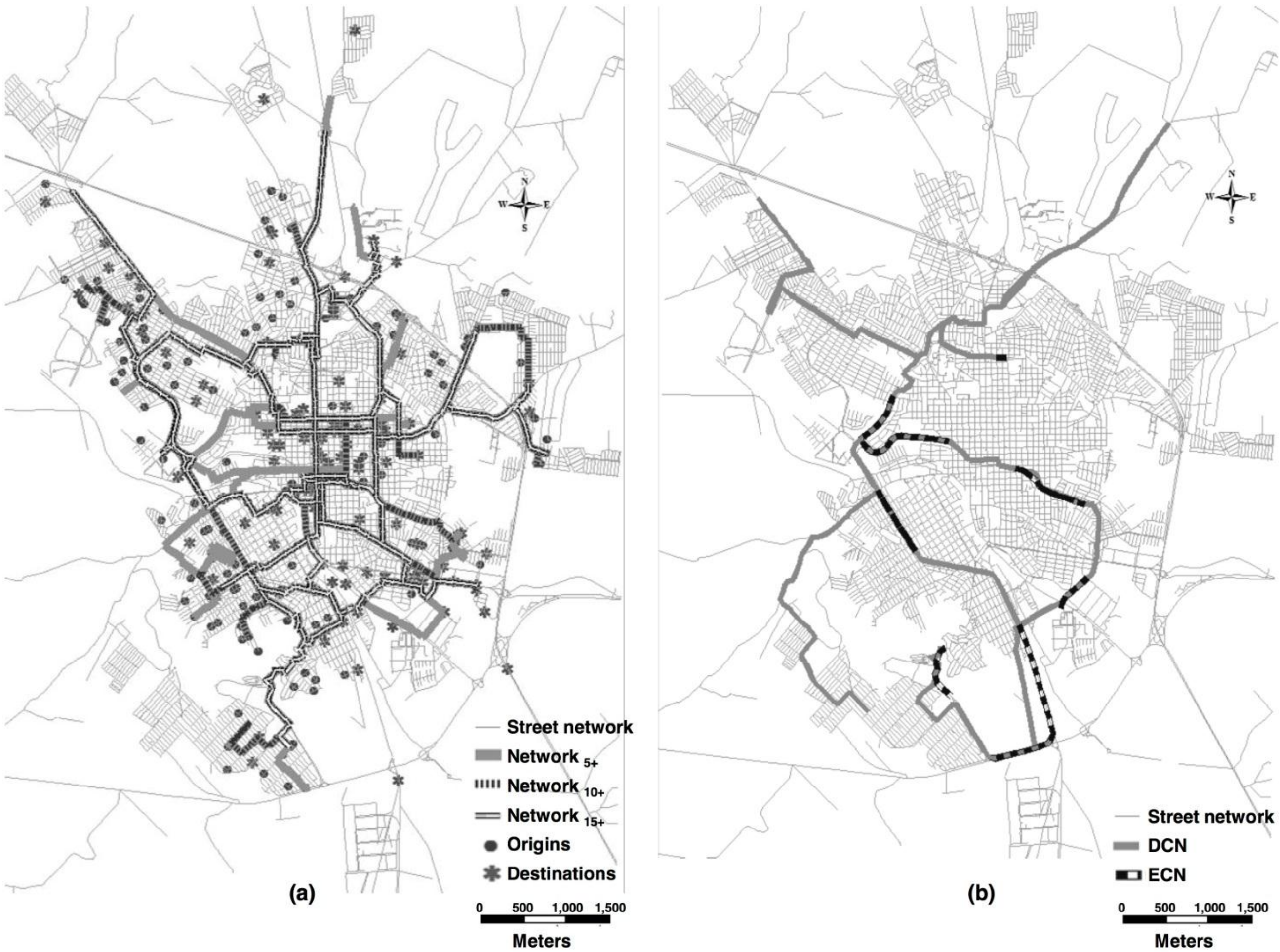

Figure 3: Cycling networks: (a) proposed by this study (b) designed by the municipality.

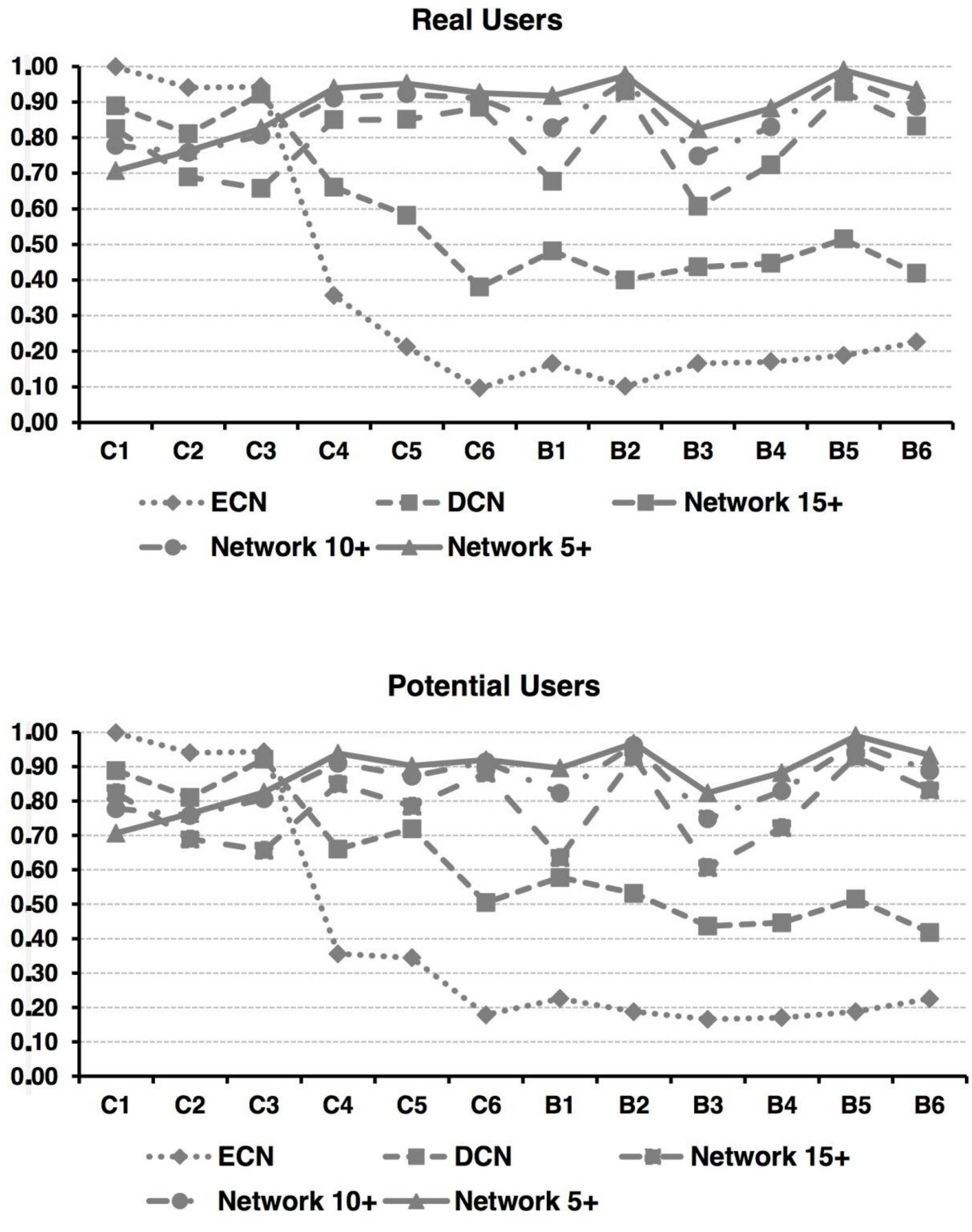

Figure 4: Normalization of cost and benefit elements considered for the cycling networks