Cycling Network Projects: A Decision-Making Aid Approach ⋆

Fernando Martínez-Plumed[1], Cèsar Ferri[1], and Lidia Contreras-Ochando[2]

[1] DSIC, Universitat Politècnica de València, Camí de Vera s/n, 46022 València, Spain.
{fmartinez,cferri}@dsic.upv.es
[2] ICube, Université de Strasbourg, CNRS, 300 Bd Sebastien Brant, BP 10413, 67412 Illkirch Cedex, France.
contrerasochando@unistra.fr

Abstract. Efficient and clean urban mobility is a key factor in quality of life and sustainability of towns and cities. Traditionally, cities have focused on cars and other fuel-based vehicles as transport means. However, several problems are directly linked to massive car use, particularly in terms of air pollution and traffic congestion. Several works reckon that vehicle emissions produce over 90% of air pollution. One way to reduce the use of fuel-based vehicles (and thus the emission of pollutants) is to create efficient, easily accessible and secure bike lane networks which, as many studies show, promote cycling as a major mean of conveyance. In this regard, this paper presents an approach to design and calculate bike lane networks based on the use of open data about the historical use of a urban bike rental services. Concretely, we model this task as a network design problem (NDP) and we study four different optimisation strategies to solve it. We test these methods using data of the city of Valencia (Spain). Our experiments conclude that an optimisation approach based on genetic programming obtains the best performance. The proposed method can be easily used to improve or extend bike lane networks based on historic bike use data in other cities.

Keywords: Bike lane network design, Optimisation, Open Data, Bike sharing, Sustainability

⋆ This work has been partially supported by the EU (FEDER) and Spanish MINECO grant TIN2015-69175-C4-1-R, and the REFRAME project, granted by the European Coordinated Research on Long-term Challenges in Information and Communication Sciences Technologies ERA-Net (CHIST-ERA), and funded by MINECO in Spain (PCIN-2013-037), by Generalitat Valenciana PROMETEOII/2015/013, and by the French National Research agency (ANR). We thank the anonymous reviewers for their comments, which have helped to improve this paper significantly. We also thank Meelis Kull, Adolfo Martínez-Usó and the rest of members of the REFRAME project for their task in collecting and providing bike rental data.

1 Introduction

Cycling has long been a part of everyday sustainable urban mobility in many countries. This trend being accentuated by the demographic changes of the second half of the 20th century: majority of citizens living in urban areas. Furthermore, over the last decade there has been a substantial growth in bike commuting thanks to, among other reasons, the bicycling infrastructures developed and the onset of bike sharing systems which have been implemented in more than a thousand cities in the world [7]. These latter systems are gaining increasing popularity in many cities as an alternative to intensive car use [8].

The advantages of cycling are undeniable: reducing car use and increasing cycling results in health benefits for commuters (and citizens in general) due to the increase of physical activity and the reduction of air pollution (particularly, green house gas emissions). To put this into figures, in [29] the authors estimated that, considering the metropolitan area of a mid-size city, replacing 40% of car trips by bike or public transport could avoid an average of (over) 65 deaths annually, this also involving a reduction of up to 200,000 tons of CO_2 annually (cars and other fuel-based transports are the main source of air pollution in cities). Additionally, the European Parliamentary Research Service [3] estimated that cycling offers significant economic advantages, at least 205 billion per year, including health care savings, reduced congestion, emissions, pollution and noise. The qualities and virtues of using the bike to commute are clear and, therefore, municipal authorities should promote their use in the cities by providing efficient and safe infrastructures.

In metropolitan areas, cycling infrastructure is generally required for commuting to work, school, or shopping. In this sense, bicycle lanes are necessary and they should be thought as efficient connections between popular origins and destinations for the commuters [10]. Bike use (or demand) as well as the effectiveness of urban bike rental services depends thus on how well-planned are the layout of stations and bike lanes. Designing this layout requires the analysis of long-term patterns in the mobility of people and thus the use of *open data* (public data and information available from government and other sources) about city mobility. For instance, there are many factors influencing bike demand, such as time of day, time of week, season, weather, current bike station geographical location, status and others events [15].

Access to public data can only be possible with the collaboration of many contributors and data providers: the public, private and non-profit sectors, as well as citizens. Furthermore, researchers, policy makers, institutions and companies are beginning to realise how important the massive amount of (open) data produced and collected is for *smart cities*. All this data can be converted into useful information for the common benefit: identify social needs, provide new services, predict and prevent disasters, etc [24]. Some examples of companies following this trend are Google Research, IBM Research, Microsoft Research [1] or *Telefónica I+D* [3].

[3] `https://ecf.com/groups/economic-benefits-cycling-eu-27`

With all of this in mind, in this paper we present (and compare) several approaches for computing comprehensive and efficient bike lanes networks at the city level. The final goal is thus improving the cycling infrastructure and aid decision-making for regional governments when planning new bicycle lanes. Through the use of open data we also aim to support cities in making the transition to "smart cities". As starting point, and in order to illustrate how our approach works, we have focused on the city of Valencia, Spain (although our approach could by applied to potentially any city with bicycling infrastructures). Valencia is the perfect example of an environmental-friendly place which is making a concerted effort to improve its bicycling: the Valencia City Council is currently studying how to improve the bike-lane network (currently having more than 180 kilometres of bike paths) with the aim of improving not only the circulation for cyclists, also their safety, while reducing traffic and the pollution levels[4]. Additionally, their citizens are showing an increasing interest in the use of the bike city: in 2014 there were 75,407 daily trips made by bicycle, 17% more than in 2009[5]. The launch of a bike sharing service (*valenbisi*[6]) provided by *JCDecaux* in 2010 has been also a turning point in the use of the bike in the city of Valencia with an average of 19,773 daily trips made by public bikes.

This document is organised as following. Section 2 summarises how we collected data used for the methods. Section 3 presents our approach, the optimisation techniques used and some experiments performed. Section 4 briefly reviews the related work. Finally, some conclusions are listed in Section 5

2 Data collection

In this section we will focus on the data acquisition, transformation an cleaning processes. Two main sources of data have been used: bike sharing demand and descriptive data about bike stations and current bike lane network. In the following we will describe them in detail.

2.1 Demand estimation

We have collected bike rental data from the 276 bike stations of the city of Valencia (Spain) provided by *JCDecaux* open data service for bike rental information[7]. Data collection started on September, 2014 and has been running continuously since then. However, the data used in our approach covers the period from January 2015 to December 2015. Bike rental data is downloaded once every minute (the average interval between two updates is 7 minutes) from all bike stations provided by the *JCDecaux* open data serviced. This service delivers the following data for each station:

[4] In `https://goo.gl/Y2SOts` we can find some of the government future plans concerning the improvement of the cycling network in Valencia (information in Spanish).

[5] See `http://www.valencia.es/ayuntamiento/estadistica.nsf` for the most recent complete statistics (in spanish)

[6] `http://www.valenbisi.com/`

[7] `http://developer.jcdecaux.com`

– **Facts of stations** (static data): station number, name, address, geographical location (latitude, longitude), number of docks (stands), availability of the station, banking facilities (pay with credit card). All these properties for one station do not change over time, although the latter three attributes can change very seldom.
– **Temporal information** (dynamic data): time of last update, the number of available bikes at that time, the number of available empty docks at that time.

Static data can be downloaded manually in text file format or accessed through the given API. Dynamic data are refreshed every minute and can be accessed only through the API by means of a personal and free API key provided by the service. Request to the dynamic data uses a GET call where one of the parameters is your API key and the user gets data in JSON format. Data are provided under an Open Data license (ODC-BY, CC-BY 2.0). There are some missing values due to denials of service from the open data server. In total, there are less than 0.7% of missing bike data. The data used is publicly available[8].

The data collected does not explicitly provide the demand of bikes (number of rented bikes per time interval) but we can estimate this number from the updates to the number of bikes in the stations: if the number of bikes in the station has decreased by k, then we know that at least k bikes have been rented in that time interval. However, this is a lower estimation of demand since it is also possible that n bikes have been returned and $n + k$ have been rented out. Another factor which affects demand estimation is the "balancing" activities carried out by the bike rental company (moving some of the bikes from one station to another). However both factors have an relative small effect in the demand (71% of times, the number of bikes decreases by exactly 1).

As we will see in the following sections, since we are interested on using demand information to weight the different bike rental stations (the bigger the bike demand is, the more important a bike rental station could be considered), and knowing that bike rental data have a strong weekly periodic component [2], we consider weekly demand profiles. To obtain these profiles, we first aggregate our data to obtain demand at every hour (calculated as the sum of decrements across that hour). Therefore, for each station we obtain a vector of length 168 where the N^{th} element quantifies the average demand in the N^{th} hour of the week (Sunday 0am-1am is the 1st hour and Saturday 11pm-12pm is the 168th hour of the week). Figure 1 shows the weekly profiles for a random selection of 10 stations.

2.2 Bike lane location

The city of Valencia has a network of interconnected cycle routes of more than 180 kilometres (this value increases every year), distributed between bike lanes, the so-called cycle-streets, and some "shared" areas with pedestrians. We have

[8] See `http://www.dsic.upv.es/~flip/BikeSharingDemand/`.

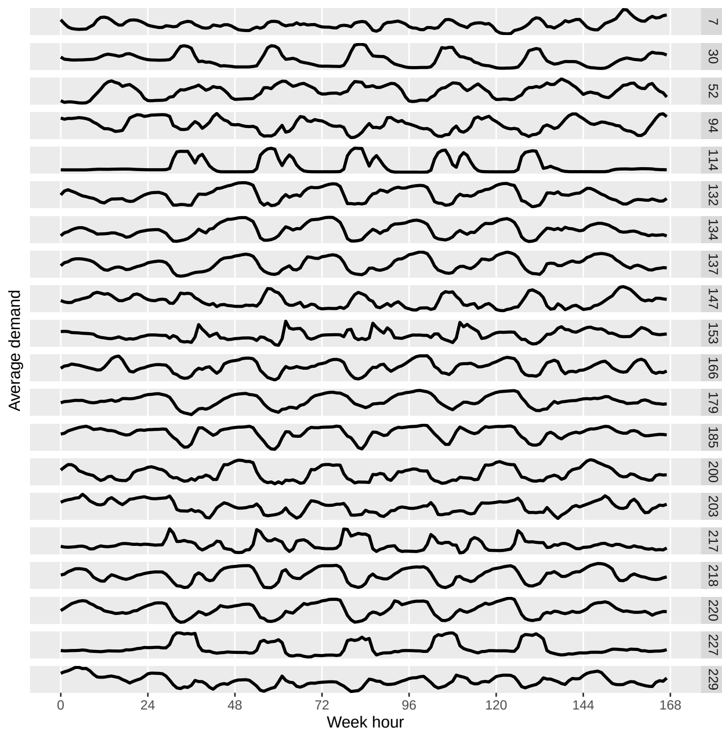

Fig. 1: Weekly profiles (starting on Sunday) of randomly selected 20 stations (*id* of the station on the right). Working days and weekend days demand patterns usually differ. The vertical scale is different for each station to fit the data.

collected location data for all the bike lanes placed in the city of Valencia. This data is easily accessible and freely available in the *Transparency and Open Data* portal[9] which provides public access to the data catalogue and information of the Valencia City Council. In particular, the data downloaded is in KML[10] format which contains information about the location of each bike lane as a vector of coordinates (latitude and longitude) that characterises the complete route. Figure 2 includes the current map of bike-lanes of the city, obtained from the aforementioned downloaded information.

[9] `http://gobiernoabierto.valencia.es/en/`

[10] *Keyhole Markup Language* (KML) is a tag-based file format with nested elements and attributes (similar to the XML standard), and it is used to display geographic data in an Earth browser such as Google Earth.

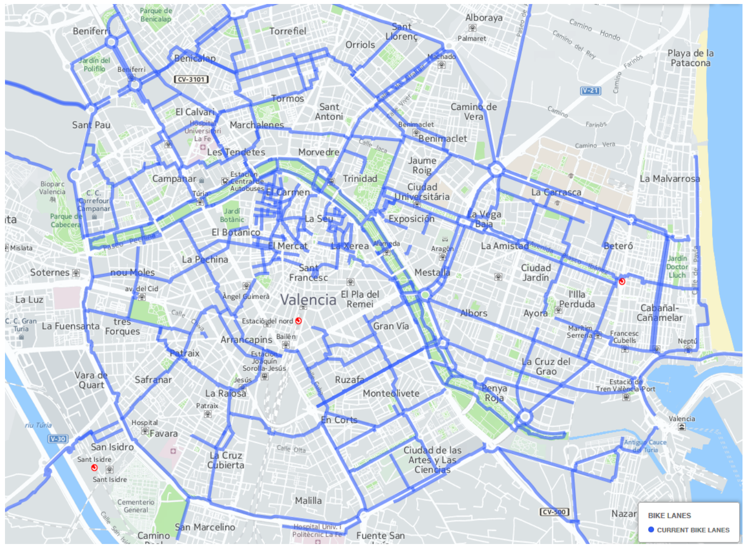

Fig. 2: Bike lanes of the city of Valencia plotted using maps provided by *Open-StreetMap* (`http://www.openstreetmap.es/`

3 Approach: bike lane network optimisation problem

As commented in the introduction, the principal aim of this work is to analyse and estimate optimal bike lane networks based on the use of open data about the historical use of a urban bike rental services. Furthermore, we want to compare the computed network solutions with the current bike lane system in order to aid the decision-making process of constructing new bike lanes. Those connections (edges in a network) returned by our approach, and not reflected in the current map of bike routes in Valencia, could be considered for construction in accordance with the stated economic and relevance criteria. For this goal, in this section we will define this task as a network design and optimisation problem and we will show some empirical experiments performed to test and compare the different strategies implemented.

3.1 Problem formulation

Given a set V of bike stations $\{v_i | i \in [1, N]\}$ where N is the number of stations ($N = 276$ *valenbisi* stations). We define D_R as distance matrix of size $N \times N$ that contains distances in kilometres between all the stations in V. This matrix is computed using the geographical location of the bike stations and applying the Haversine formula [32]. This formula provides great-circle distances between two points on a sphere from their longitudes and latitudes, that is, it could be used to obtain the shortest distance between two bike stations over the earth's surface given its radius in kilometres[11]. For any two points on a sphere, the Haversine of the central angle between them is given by

$$hav(\frac{d}{r}) = hav(\varphi_2 - \varphi_1) + cos(\varphi_1) \cdot cos(\varphi_2) \cdot hav(\lambda_2 - \lambda_1)$$

where $hav(\theta) = sin^2(\frac{\theta}{2}) = \frac{1-cos(\theta)}{2}$, d is the distance between the two points, r is the radius of the sphere, and $(\varphi_1, \lambda_1), (\varphi_2, \lambda_2)$ are the coordinates of the points (latitude, longitude). For the implementation of this formula, we use the R package `geosphere` [19].

We define a solution S as a Boolean matrix of size $N \times N$. A $true$ value in the matrix position i, j indicates that the solution includes a direct bike path connection between stations i and j. Note that, to avoid duplicated connections, we only consider solutions where $S_{i,j} = false$ if $j \geq i$. Given V and S, we generate an undirected graph $G(V, S)$ where there is an edge between stations i and j if $S_{i,j} = true$ (for practical purpouses $true \equiv 1$ and $false \equiv 0$) . Given a solution S, we define the length of a solution such as $L(S) = \sum_{i=1}^{N} \sum_{j=1}^{N} (S_{i,j} \cdot D_{R(i,j)})$, namely, the total length (in kilometres) of the proposed bike lane network. We also define the constant value L_{max} as the maximum permitted length (also in kilometres) of a solution network.

Regarding the number of possible valid solutions (matrices $S \in \mathcal{S}$), there are $2^{N(N-1)/2}$ different matrices. However, not all of these solutions are useful. In order for a solution S to be valid, the following constraints have to be met:

1. All the stations must be connected, i.e., there is only one fully connected network component in $G(V, S)$.
2. The length of the solution networks must not be grater than L_{max}, i.e. $L(s) \leq L_{max}$.

Furthermore, not all the solutions are equally interesting. Although we could use the length of the solution as a measure of performance (the shorter the network is, the better the solution is), we are more interested in increasing the probabilities of connecting (directly) those more relevant (most used) stations. For that reason we define w as an array of size N, where $\{w_i | i \in [1, N]\}$ represents the importance (or weight) of station i (values between $[0, 1]$). These weights are computed taking into account the historical demand or use of each station in the network V. Additionally, we define W as a matrix of size $N \times N$ such that $W_{i,j} = w_i \cdot w_j$. This matrix represents the importance of every pair of possible stations. Additionally, we define $D_G(S)$ as distance matrix of size $N \times N$ that contains distances in kilometres between all the stations in V. In other words, $D_G(S)$ contains the length in kilometres for each pair of stations using the shortest path in the graph $G(V, S)$ between them. By definition, $D_G(S)_{i,j} = 0$ if $j \geq i$. We use Dijkstra's algorithm to compute the distance (shortest path) between stations. Finally, as a measure of optimality, we define the cost C of

[11] The alternative would be to use real distances computed considering the path in the streets of the city between stations.

a solution S as $C(S) = \sum_{i=1}^{N} \sum_{j=1}^{N} (W_{i,j} \cdot D_{G(i,j)}(S))$. In this way, we use as cost of a solution the sum of the distances of travelling for each pair of stations using $G(V, S)$ weighted by the importance of the stations. Therefore, we want to find solutions that minimise bike path distances between stations in V giving more relevance to the most used stations. Summing up, given a set of stations V, a matrix of distance D_R between the stations in V and an array of stations weights w defining their relevance, the goal of our approach is to find a valid solution S with the minimum $C(S)$.

Last but not least, we need to define the weights that characterise each bike station (w_i). These weights seek to reflect not only historical demand (number of rented bikes) of the bike stations (average usage during 2015), also their size in terms of number of docks (stands) (the more docks they have, the more relevant they can be considered). As we have seen in the previous section, bike rental data have a strong weekly periodic component, so we have considered weekly demand profiles for each station. Furthermore, it is easy to see (Figure 1) that the demand per station usually follow a different pattern (regularity) during working days and during weekend days respectively, i.e., each station have thus two clear distinct patterns. Therefore, as a very simple approach, we characterise the average demand for each station as the weighted sum of *(a)* the average working days demand plus *(b)* the average weekend days demand of its weekly profile. Furthermore, we take into account the importance of the stations in terms of their size (the number of docks varies from 10 to 40), so we finally characterise an station as the weighted sum of both previous terms (all values are normalised between 0 and 1). Figure 3 illustrates the weight associated to each *valenbisi* bike station in the city of Valencia.

3.2 Experiments

In this part we include the results of the experiments performed. We test several optimisation strategies aiming at finding a suboptimal solution to the network configuration problem defined in the previous section. In particular, as a preliminary approach for exploring the solution space we use four common and well-known randomised search methods. These optimisation methods introduce randomness into the search-process to accelerate progress and thus it is possible to escape a local optimum and eventually to approach a global optimum [20]. The optimisation methods used are briefly described in the following:

- **Monte Carlo**: Monte Carlo methods are a broad class of computational algorithms that approximates solutions to quantitative problems through repeated statistical sampling. The key idea is to use randomness in order to address problems that might be deterministic in principle and too complicated to be solved analytically. For the problem at hand, we randomly search the space of solutions by creating a large number of possible *valid* solutions (e.g., 1000 possible different bike lane networks) and then we select the best one (the solution with the lowest computed cost).

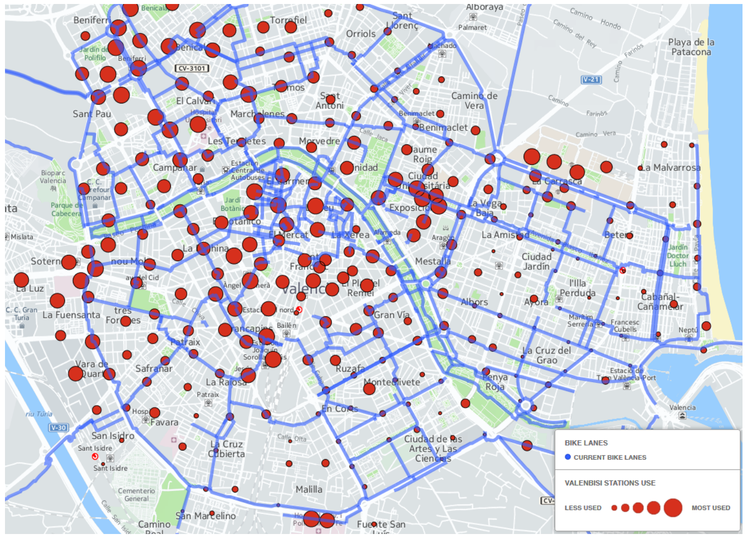

Fig. 3: Weight associated to each bike station plotted using maps of CartoDB platform (`https://cartodb.com/`). The bigger the red point, the more demanded has been the bike station.

- **HillClimb**: The hill climbing optimisation technique is an iterative algorithm that, for a given problem, starts with the selection of an arbitrary solution and then attempts to find a better solution by exploring the neighbouring solutions (incrementally changing a single element of the former considered solution). In our case, we start from a valid solution and then we create k neighbouring solutions. Each new solution is created by randomly adding a new connection (a $True$ value in S) for each of the N rows in S. If the change produces a better solution (solution with a lower cost), an incremental change is made to the new solution, repeating until no further improvements can be found.This process can be seen as going down a hill.
- **Simulated Annealing**: This technique is inspired by physics processes. Annealing is a softening process in which metals are heated and then allowed to cool slowly. The atoms are first highly excited and then gradually settle into a low energy state. In this way the atoms are able to find a stable configuration. In the same way as the previous techniques, our simulated annealing algorithm starts with a random solution. In each iteration, we slightly modify the current solution by randomly adding or removing a connection between stations. The key difference with the above methods is that it uses a parameter called *temperature* used for calculating the acceptance probability of inferior neighbouring solutions (candidate solutions with higher cost). This parameter usually starts at 1.0 and is decreased at the end of each iteration by multiplying it by a *cooling rate* constant. The process works this way: if the cost of a candidate solution is lower than the cost of current solution, it is

accepted as the new solution. If the cost is higher, the candidate can still be accepted according to a given probability: the higher the temperature value is, the higher the odds are for the solution to be accepted. The algorithm stops when the temperature parameter reaches a bottom limit.

– **Genetic Algorithm**: This family of techniques are based on adaptive heuristic search based on the evolutionary ideas and processes of natural selection and genetics. Genetic algorithms repeatedly modify a population of individual solutions by using different techniques of natural evolution: inheritance, mutation, selection and crossover. In our implementation, at each step, the genetic algorithm selects individuals ("elite" set formed by the ten best solutions) from the current population (starting with 100 valid solutions) to be parents and uses them to produce the children for the next generation. Over successive generations, the population *evolves* toward an optimal solution. We perform 300 iterations in order to obtain the optimal solution.

All these previous optimisation techniques have been implemented in R[12]. We use package `igraph`[13] which provides functions for generating, visualising and handling large scale networks and graphs. The source code and data used for the experiments can be found in GitHub[14]

α	20			30			40			50		
Method	Cost	Length	Time	Cost	Length	Time	Cost	Length	Time	Cost	Length	Time
Montecarlo	287	1927	694	287	1927	707	287	1927	613	287	1927	634
HillClimb	32	5919	85	40	3946	41	47	2959	22	54	2367	13
Anneal	30	5920	330	34	3947	273	40	2960	236	46	2368	223
Genetic	**25**	5916	544	**29**	3941	370	**34**	2956	313	**39**	2362	280

Table 1: Average results (10 repetitions) for each optimisation method and network limit ($\alpha = [20, 30, 40, 50]$). Table shows the average cost of the solution (*Cost*), the average length in kilometres of the proposed solution (*Length*) and the average time in seconds required to compute the solution (*Time*). Bold values represent the best solutions (cost-effective).

In Table 1 we include the results of the experiments using the data from the city of Valencia (described in Section 2) and the four optimisation techniques previously introduced. In this table we show the results for each method tested under four different scenarios. In each scenario we consider a different upper limit length (in kilometres) for the bike lane network to be estimated. We define the L_{total} constant as the total length of this network, on the assumption that all *valenbisi* stations are directly connected (118,411Km for the current network).

[12] See `http://caret.r-forge.r-project.org`.

[13] See `https://cran.r-project.org/web/packages/igraph/index.html`

[14] For reproducibility, all the experiments and the source code can be found in`https://github.com/ceferra/ValenbisiNetwork`

We already defined L_{max} as the length maximum of a solution in the network of bike lanes. Then, we explore four different L_{max} values (for illustrative purposes), $L_{max} = L_{total}/\alpha$ where $\alpha = [20, 30, 40, 50]$, namely, the greater the α value is, the smaller the L_{max} value is. The table contains the average results (10 repetitions) for each approach and scenario. We show three results for each method. The *cost* of the solution ($C(S)$), the *average time* in seconds required to compute the solution, and the *size* in kilometres of the proposed solution. The cost is computed as defined previously.

From the table we see that, when we increase the value of the factor α, we (obviously) limit the size of the valid solutions. However, this reduction in the size of the estimated network is reflected in higher costs since we are removing direct connections (edges) between stations. If we compare the four optimisation techniques (attending to the cost of the solution), the same results are obtained for all the scenarios studied: the best results are obtained by the the genetic-based technique, followed by simulated annealing and in third position comes Hillclimb. As expected, Montecarlo gets always the worst performance due to its random nature. It is noticeable that, since Montecarlo selects costly solutions but small networks at the same time, we obtain the same results for the different values of *alpha*. Regarding the time values, we see that Hillclimb obtains the best performance due to its greedy behaviour. Figure 4 shows an illustrative example of an estimated network solution by using the genetic approach and the biggest α value (for visualisation reasons we use just the 25 most relevant bike stations).

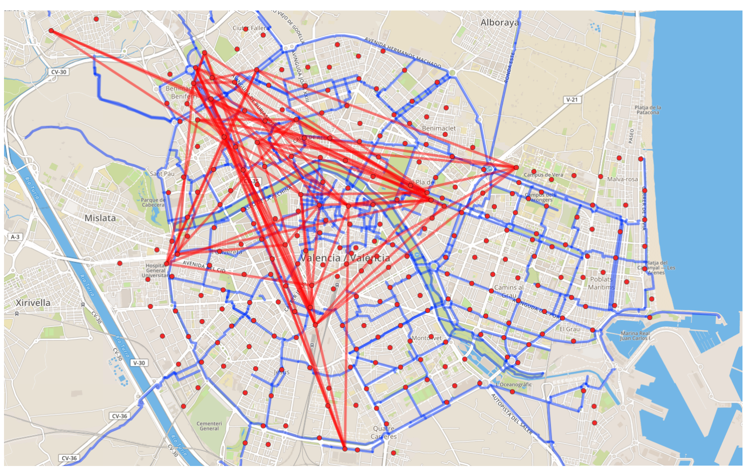

Fig. 4: Obtained solution by using the genetic algorithm approach. Due to visualisation reasons, we only use the 25 most relevant bike stations (according to its historical use).

Considering the limited scope of the open data employed in the optimisation methods, we need to remark that results in Table 1 are just a simple approach to the the network design problem (NDP) [21]. Additionally, the resulting networks are not realistic (regarding their length) due to the fact that we only consider straight line connections between stations. However, we believe that fine-tuning our method and techniques, using exact distances between stations, taking into account further contextual data, and performing further experiments, we could get better and realistic results. In any case, the estimated bike lane networks can be useful to discover which new connections might be added to the current network, according to criteria such as historical use and importance of the bike stations.

Considering the best cost-effective solution (genetic approach) and analysing the current bike lane network in the city of Valencia, we are able to propose the construction of some new bike lanes aiming at improving the current network (and purely for illustrative purposes). We show the estimated solution in the left part of Figure 5. This map shows the current network of bike lanes in solid blue lines and the new proposed ones are drawn as dashed black lines. Recently, the city of Valencia announced an extension of the bike lane network. In the right section of Figure 5 we can see the proposed extension. New bike lanes are shown as dashed red lines. As we can see, there are relevant coincidences between both proposals (highlighted in yellow). These coincidences show that our proposal provide a useful aid for extending bike lane networks (it is possible to provide a greater number of connections changing the maximum length requirements).

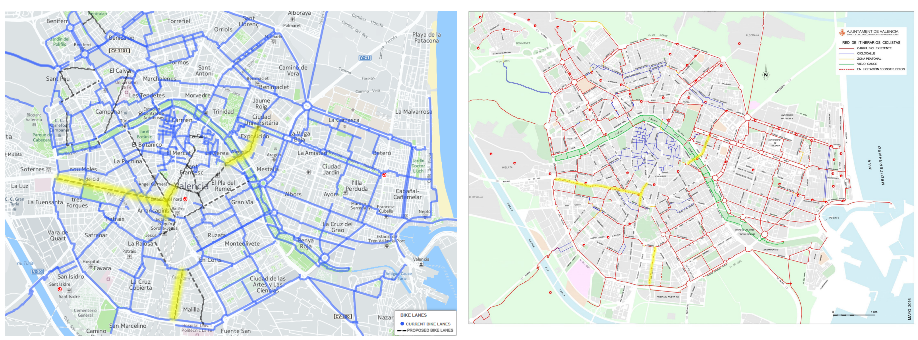

Fig. 5: Left: Extension of the bike lane network for the city of Valencia based on the solution in Figure 4. New lines are shown as dashed black lines. Right: Official planned extension of the bike lane network in Valencia, new lanes are the dashed red lines. Coincidences between both extensions are highlighted in yellow.

Finally, we have created a public website[15] with the objective of providing information about the estimated bike lane networks (graphs and features about *valenbisi* stations). Written in `PHP` and `jQuery`, and using `javascript` libraries

[15] `http://users.dsic.upv.es/~flip/RutasBici/`

such as `leaflet`[16] as well as plugins and services of `Mapbox` platform[17], the website uses `OpenStreetMap` maps with different interactive layers showing different data: *(1)* the geographical location of all *valenbisi* stations, *(2)* the geographical location of all the bike lanes in the city of Valencia, and *(3)* the optimal solution (using the genetic approach) for connecting the 276 *Valenbisi* stations.

4 Related work

The stated problem of building new bike lane connections in an existent bike lane network is closely related to the network design problem (NDP) [21]. This latter is defined as, given a weighted graph, we want to select a subgraph that satisfies the transportation demand and minimises the overall costs of the transportation. This is one of the most challenging transport problems and it has usually been addressed in two different ways: exact solutions [11,16,4], which deal with NDP in a rigorous but inefficient manner; and heuristic solutions [12,22,14,26] which provide approximate yet efficient solutions even for large-scale real-world networks. The latter approaches are more popular than exact solutions.

In particular, our approach was initially inspired by the ideas from [34] which proposes a biologically-based adaptive network design. In particular, they applied a mathematical model of the behaviour of the a cellular slime mould *Physarum polycephalum* [33] (a unique creature which produces efficient networks when foraging for its food), to solve a NDP by "maximising transport capacity of the network and minimising the size and length of the network". In particular, in [34] the authors showed how the slime mould created a network similar to the existing Tokyo train system (oat flakes were dispersed to represent Tokyo and 36 surrounding towns), and "with comparable efficiency, fault tolerance, and cost".

Explicitly focusing on the problem of improving the cycling network in a city, most of the literature concentrates on the estimation of the potential demand to the services [5,36,17,13,27] as well as solving the tendency of (these stations) become quickly unbalanced, with some stations being mostly used to pick up bicycles and other more to return bicycles [30,31,28,9,6,23].

However, there are much less approaches related to the decision-making process for constructing bike lanes. Some examples include [18] which presents a wide analysis for the improvement of the bike infrastructures in Columbus (USA). This work shows a recommended bicycle network for the city based on the study and analysis of the needs and types of bicyclists, an estimating of the existing demand, air quality benefits, etc. With this information, the authors present a blueprint for how the city can accommodate, plan for, and promote bicycling with a new network, parking facilities and maintenance. We find another example in [25] where the authors address the strategic planning of public

[16] Leaflet is an open-source JavaScript library for mobile-friendly interactive maps. See `http://leafletjs.com/`

[17] Mapbox is a full-featured design suite for creating custom map styles. See `https://www.mapbox.com/`

bicycle sharing system (location of bike stations and network structure) by using mathematical models which integrates the travel costs of users, the facility costs of bike stations, the setup costs of bicycle lanes, as well as the service level. More recently, [35] presents a mixed-integer programming formulation for designing the service network in metropolitan areas. Their model also anticipates operational relocation decisions by a dynamic transportation model yielding relocation services.

5 Conclusion

Sustainable urban mobility requires a gradual increase in the use of bikes as a major mean of conveyance. An efficient and well-designed bike lane network is a relevant factor for promoting an everyday use of the bike for a variety of reasons. For instance, safety is significantly increased since the existence of a bike lane network helps to reduce the rate of cyclist accidents. Furthermore, some studies have shown that the bike lines stimulate the local economy as they facilitate commuting. Finally, bike lines have a real impact on the environment since cycling reduces ones carbon footprint by not burning fossil fuels, thus avoiding the emission associated pollutants.

In this paper we have presented an approach for computing and analysing bike lane networks from open data of bike sharing systems. Specifically, we use bike station locations and their historical use (demand) in order to characterise each station independently. With all this information, we address the task as a network design problem (NDP) where stations are seen as nodes and bike lanes are considered as edges of the network. Four different optimisation strategies have been studied and analysed: Montecarlo, HillClimb, simulated annealing and genetic algorithm. The experiments using open data from the city of Valencia (Spain) show that the technique based on the genetic algorithm obtains the best performance. We have compared the obtained solution with the current bike lane network in the city of Valencia, obtaining several new bike lanes which could be useful to extend the existing network. It is worth highlighting that some of these new lines correspond with some of the new bike lanes that the city of Valencia has already scheduled to be assembled.

We may emphasise once again the preliminary character of our approach. Therefore, based on our own experimentation and the related bibliography, we plan to explore and analyse new optimisation strategies and techniques (such as SAT based solvers for optimization problems), including a further improvement of the current methods.

References

1. Rajendra Akerkar. *Big data computing*. CRC Press, 2013.
2. Lars Böcker, Martin Dijst, and Jan Prillwitz. Impact of everyday weather on individual daily travel behaviours in perspective: a literature review. *Transport reviews*, 33(1):71–91, 2013.

3. Andrey Bogomolov, Bruno Lepri, Jacopo Staiano, Nuria Oliver, Fabio Pianesi, and Alex Pentland. Once upon a crime: towards crime prediction from demographics and mobile data. In *Proceedings of the 16th international conference on multimodal interaction*, pages 427–434. ACM, 2014.
4. Gary A Davis. Exact local solution of the continuous network design problem via stochastic user equilibrium assignment. *Transportation Research Part B: Methodological*, 28(1):61–75, 1994.
5. Juan de Dios Ortuzar, Andres Iacobelli, and Claudio Valeze. Estimating demand for a cycle-way network. *Transportation Research Part A: Policy and Practice*, 34(5):353–373, 2000.
6. Mauro Dell'Amico, Eleni Hadjicostantinou, Manuel Iori, and Stefano Novellani. The bike sharing rebalancing problem: Mathematical formulations and benchmark instances. *Omega*, 45:7–19, 2014.
7. P. DeMaio and R Meddin. The bike-sharing world map, 2016.
8. Paul DeMaio. Bike-sharing: History, impacts, models of provision, and future. *Journal of Public Transportation*, 12(4):3, 2009.
9. Luca Di Gaspero, Andrea Rendl, and Tommaso Urli. Constraint-based approaches for balancing bike sharing systems. In *International Conference on Principles and Practice of Constraint Programming*, pages 758–773. Springer, 2013.
10. Carr T. Dill J. Bicycle commuting and facilities in major u.s. cities: If you build them, commuters will use them - another look. *Transportation Research Record: Journal of the Transportation Research Board*, 1828, 2003.
11. René Dionne and Michael Florian. Exact and approximate algorithms for optimal network design. *Networks*, 9(1):37–59, 1979.
12. Falko Dressler and Ozgur B Akan. Bio-inspired networking: from theory to practice. *IEEE Communications Magazine*, 48(11):176–183, 2010.
13. Côme Etienne and Oukhellou Latifa. Model-based count series clustering for bike sharing system usage mining: A case study with the vélibsystem of paris. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3):39, 2014.
14. Wei Fan and Randy B Machemehl. Optimal transit route network design problem with variable transit demand: genetic algorithm approach. *Journal of transportation engineering*, 132(1):40–51, 2006.
15. Hadi Fanaee-T and Joao Gama. Event labeling combining ensemble detectors and background knowledge. *Progress in Artificial Intelligence*, 2(2-3):113–127, 2014.
16. Virginie Gabrel, Arnaud Knippel, and Michel Minoux. Exact solution of multicommodity network optimization problems with general step cost functions. *Operations Research Letters*, 25(1):15–23, 1999.
17. Juan Carlos García-Palomares, Javier Gutiérrez, and Marta Latorre. Optimizing the location of stations in bike-sharing programs: a gis approach. *Applied Geography*, 35(1):235–246, 2012.
18. Columbus Government. Columbus bicentennial bikeways plan. `http://nacto.org/wp-content/uploads/2012/07/ColumbusBMPFinalApril2008.pdf`, 2008.
19. Robert J Hijmans, Ed Williams, Chris Vennes, and Maintainer Robert J Hijmans. Package geosphere, 2015.
20. Holger H Hoos and Thomas Stützle. *Stochastic local search: Foundations & applications*. Elsevier, 2004.
21. David S Johnson, Jan Karel Lenstra, and AHG Kan. The complexity of the network design problem. *Networks*, 8(4):279–285, 1978.
22. Chia-Feng Juang. A hybrid of genetic algorithm and particle swarm optimization for recurrent network design. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(2):997–1006, 2004.

23. Christian Kloimüllner, Petrina Papazek, Bin Hu, and Günther R Raidl. Balancing bicycle sharing systems: an approach for the dynamic case. In *European Conference on Evolutionary Computation in Combinatorial Optimization*, pages 73–84. Springer, 2014.
24. John Lim and Daniel Yeap. What is social innovation? 2012.
25. Jenn-Rong Lin and Ta-Hui Yang. Strategic design of public bicycle sharing systems with service level constraints. *Transportation Research Part E: Logistics and Transportation Review*, 47(2):284 – 294, 2011.
26. Chi-Chun Lo and Wei-Hsin Chang. A multiobjective hybrid genetic algorithm for the capacitated multipoint network design problem. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 30(3):461–470, 2000.
27. REFRAME Project. Morebikes: 2015 ecml-pkdd challenge on model reuse with bike rental station data, 2015.
28. Tal Raviv, Michal Tzur, and Iris A Forma. Static repositioning in a bike-sharing system: models and solution approaches. *EURO Journal on Transportation and Logistics*, 2(3):187–229, 2013.
29. D. Rojas-Rueda, A. de Nazelle, O. Teixid, and M.J. Nieuwenhuijsen. Replacing car trips by increasing bike and public transport in the greater barcelona metropolitan area: A health impact assessment study. *Environment International*, 49:100 – 109, 2012.
30. Arieh Schlote, Bei Chen, Mathieu Sinn, and Robert Shorten. The effect of feedback in the assignment problem in shared bicycle systems. In *2013 International Conference on Connected Vehicles and Expo (ICCVE)*, 2013.
31. Jasper Schuijbroek, Robert Hampshire, and Willem-Jan van Hoeve. Inventory rebalancing and vehicle routing in bike sharing systems. 2013.
32. BP Shumaker and RW Sinnott. Astronomical computing: 1. computing under the open sky. 2. virtues of the haversine. *Sky and telescope*, 68:158–159, 1984.
33. Atsushi Tero, Ryo Kobayashi, and Toshiyuki Nakagaki. A mathematical model for adaptive transport network in path finding by true slime mold. *Journal of theoretical biology*, 244(4):553–564, 2007.
34. Atsushi Tero, Seiji Takagi, Tetsu Saigusa, Kentaro Ito, Dan P Bebber, Mark D Fricker, Kenji Yumiki, Ryo Kobayashi, and Toshiyuki Nakagaki. Rules for biologically inspired adaptive network design. *Science*, 327(5964):439–442, 2010.
35. Patrick Vogel. Service network design of bike sharing systems. In *Service Network Design of Bike Sharing Systems*, pages 113–135. Springer, 2016.
36. Patrick Vogel and Dirk C Mattfeld. Modeling of repositioning activities in bike-sharing systems. In *World Conference on Transport Research (WCTR)*, 2010.