Noname manuscript No.
(will be inserted by the editor)

Using the districting problem in Bicycle Sharing Systems to facilitate the balancing of the operation

Pablo A. Maya Duque · Diana M. Peréz Valencia · Miguel A. Arroyave Guerrero

Received: date / Accepted: date

Abstract The Bicycle Sharing Systems (BSS) offer a mobility service in which public bicycles are available for shared use. The demand of this type of systems is characterized to be unpredictable, asymmetric and spatial-time dependant. These demand characteristics affect the system balance during specific periods of time. That is, bicycles accumulate in some stations, leaving no free parking docks for incoming users, while other stations are empty not being able to satisfy new users demand. The repositioning of bicycles is the most used strategy to balance the system. In that strategy, external vehicles transport bicycles from crowded stations to empty stations in which a demand peak is foreseen. Usually, the operational area is divided into zones to be served by different repositioning vehicles. This paper addresses the districting problem arisen when creating the repositionig zones. It deals with a tactical decision (i.e., districting problem), unlike most of the research on repositioning bicycles in BSS, which focuses on operational decisions such as routing and inventory management. Moreover, this work main contribution is to take into account when defining the districts of the BSS, not only distance and connectivity, but also criteria such as demand patterns and stations criticality. A mathematical model that involves those criteria is proposed. It is tested on instances built from real operational data of ECOBICI in Mexico City, which allows to draw insights to be taken into account by the system operators.

Keywords Bicycle Sharing System; Districting problem; Repositioning strategy

P.A. Maya Duque
University of Antioquia
E-mail: pmayaduque@gmail.com

D.M. Peréz Valencia
Basque Centre for Applied Mathematics
E-mail: dimapeva@gmail.com

M.A. Arroyave Guerrero
University of Antioquia
E-mail: miguel.arroyave@udea.edu.co

1 Introduction

Bicycle Sharing Systems (BSS) have positioned as a sustainable alternative for urban mobility problematics. The Bicycle Sharing Systems offer a mobility service in which public bicycles are available for shared use. These bicycles are located at stations that are displayed across an urban area. The users of the system can take a bicycle from a station, use it for a journey, leave it in a station (not necessarily the one of departure), and then pay according to the time of usage (Dell'Amico et al., 2014).

The project OBIS, Optimising Bike Sharing in European Cities (Büttner and Petersen, 2011), identifies key influencing factors on the outcomes of BSSs that can be distinguished into endogenous and exogenous. The former are factors specific to the city and not easily changed, while the latter are policy sensitive design factors that can be adjusted. Within the exogenous factors are those related to the physical design such as service design and technology usage. Additionally, the study pointed out success factors for bike sharing schemes highlighting the importance of user accessibility, bike and station design, financing model and traffic redistribution. Most of the factors that determine the success of a BSS are related to the service level, which can be understood as the likelihood of satisfying the user need for a bike or parking dock on a specific period of time. An user would not be provided a satisfactory service level when attempting to rent a bike from an empty station or return a bike in a full station.

Among the factors that affect the service level of a BSS, one that stands out is the fact that the demand of this type of systems is characterized to be unpredictable, asymmetric and spatial-time dependant. Moreover, factors such as weather and topographical conditions could also influence the demand patterns (Faghih-Imani et al., 2014). These demand characteristics and usage patterns affect the system balance during specific periods of time. That is, bicycles accumulate in some stations, leaving no free parking docks for incoming user, while other stations are completely empty not being able to satisfy new user service demands and, therefore, affecting the service level of the BSS. Several strategies have been proposed to deal with the problem of balancing the operation of the BSS, being the repositioning of bicycles the most used in practice. In that strategy, external vehicles, usually trucks with capacity for several bicycles, transport the units from crowded stations to empty stations in which a demand peak is foreseen.

During the International meeting of Bicycle Sharing Systems, held in Medelln in 2016, we had the opportunity to learn from the experience of different BSS operators. We contacted 11 BSSs and were able to perform depth interviews with four operators responsible of five systems, namely, EnCicla (Colombia), BikeSantiago (Chile), ECOBICI (Mexico), BikeRio (Brazil) and ECOBICI (Argentina). Those interviews showed that the operators address the balancing of the system, and particularly the repositioning of bicycles, dividing the system into zones. Generally, the number of zones depends on the size and type of the fleet of repositioning vehicles. The definition of zones aims at distributing the repositioning workload among the vehicles and at the same time facilitating the routing of the fleet, particularly under heavily congestion conditions.

This paper addresses the districting problem faced by BSS operators when they have to divide the operation area of the system in a set of zones to be covered by each of the repositioning vehicles. It deals with a tactical decision (i.e., districting problem), unlike most of the research on repositioning bicycles in BSS, which focuses on operational decisions such as routing and inventory management problems. Moreover, this work main contribution is to take into account not only distance and connectivity when defining the districts of the BSS, but also criteria such as demand patterns and stations hierarchy. A mathematical model that involves those criteria is proposed, and it is tested on instances built from real operational data of ECOBICI in Mexico City.

The rest of the paper is organized as follows. Section 2 describes the problem. Section 3 contains a brief literature review on strategies to repositioning bicycles in BSS that involve tactical or strategic decisions. Section 4 presents the mathematical model and details of the proposed method. Section 5 describes the case study and presents the results and insights obtained from the experiments. Finally, Section 6 includes the concluding remarks.

2 Problem definition

The districts of a BSS, also denoted as repositioning zones, are usually created taking into account several geographical criteria being distance and connectivity the most commonly used. That is, two stations that belong to the same zone are close to each other and its is possible to travel between each pair of stations. However, as observed in some of the studied systems, limiting the definition of zones to only geographical criteria has some drawbacks. For instance, it generates that the repositioning vehicles have to travel to adjacent zones in order to be able to fulfill the requirements of the stations in the zone to which the vehicle is assigned. This is due to the fact that in specific hours, particularly peak hours, most of the stations of a zone demand bicycles (or parking docks) while only few of the stations have excess of them, therefore there is not enough supply of bicycles (or parking docks) within the zone to fulfill the aggregated demand. Figure 1 shows the real-time distribution of bicycles in a peak hour for the case of EnCicla, in Medellín Colombia. It shows that for Zone 4 (Zona 4) most of the stations are demanding bicycles but only few of the station have a slack that could be used in the repositioning operation. Additionally, there is not an even distribution of the most important stations among the zones. Therefore, if something goes wrong with the repositioning operations in an specific zone that has several important stations, the impact on the performance of system is significant. For these reasons different operators have recognized the need of considering additional criteria when solving districting problems for BSS. Some of the criteria to be considered are station requirements (i.e., demand patterns of bicycles and parking docks), priority of the stations in terms of the number of transactions and operational criticality, and vehicular congestion.

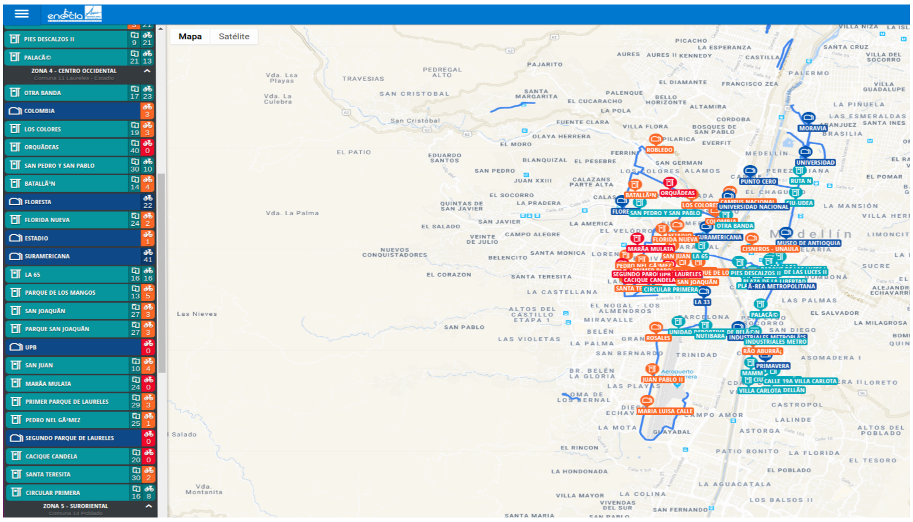


Fig. 1 Inventory distribution in a peak hour for the system EnCicla

The districting problem of a BSS divides the system into zones. Each zone is formed by a set of stations that are close to each other and connected. Thus, a set of geolocated points have to be partitioned into subsets of them that would be assigned to a repositioning vehicle. The problem differs from the districting problem applied to polygons or areas (e.g., political districting problem (Salazar-Aguilar et al., 2011)) as the basic units are points instead of polygons which, for instance, allows intersections between the perimeters or convex hulls of two different zones. Additionally, special data analysis and modeling approaches have to be consider to involve demand patterns and stations criticality within the districting decision making.

System operators and researchers have recognized the need of studying the balancing problem from a broader perspective in which strategic and tactical decisions are taken into consideration. This paper addresses that gap by considering the relation between the districting problem and the repositioning problem. It proposes a mathematical model that supports the districting of a BSS network into zones to be covered by repositioning vehicles while taking into account the demand patterns and stations importance. The model would help to understand the effect of considering those additional criteria.

3 Literature review

Based on the literature, Arroyave (2016) proposes a framework that classifies the problematics faced on the operation of a BSS considering two different perspectives: the level of planning (i.e., strategic, tactical, and operational) and the type of decisions that it involves (i.e., system design, demand management, and resource management). Figure 2 summarizes the identified problematics.

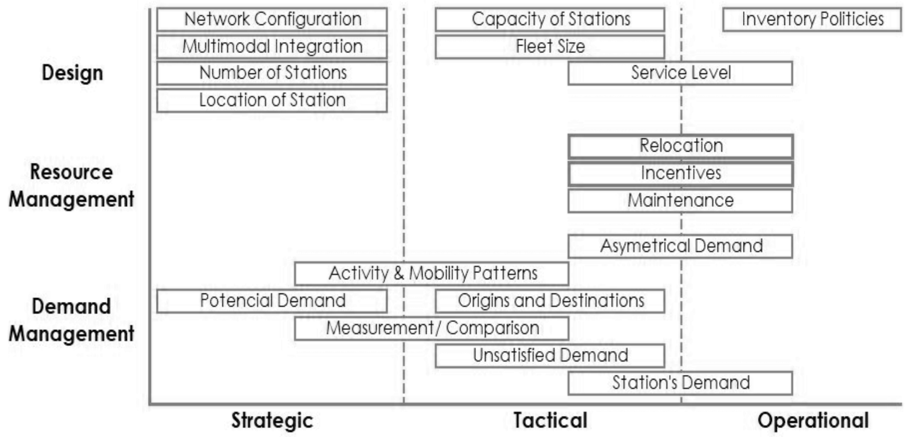


Fig. 2 Operation management problematics of the BSS.

Most of the problematics related to the BSS balancing can be found within the tactical and operative planning level and are associated to the demand and resource management decisions. The strategies to address the balancing can be classified depending on whether they act over the demand process or the supply process, although mixed approaches might be possible as well. The decisions related to the strategies that affect the supply process are usually under the system operator domain, which facilitates their implementation. Among those strategies, repositioning bicycles has been the most studied. It is usually understood as an operational task that has to be coordinated and executed in the day-to-day operation.

The repositioning can be static or dynamic (Laporte et al., 2015). The former is carried on during the night, the system idle time or periods of low demand and prepare the system for the beginning of the operation or to face peak demand periods. The dynamic repositioning is performed during the operation responding to the real-time needs of the system (Dell'Amico et al., 2014). Additionally, the repositioning of bicycles can be carried on exclusively by the system operator or can involve the users. The latter is denoted as collaborative repositioning. Most of the literature focuses on the static repositioning, performed by the system operator, involving routing decisions concerning the vehicles, and inventory decisions concerning the number of bicycles in the stations. This problematic has attracted the attention of the research community in the last decade with the number of publications and conference risen considerably in the last couple of years (Laporte et al., 2015). However, the scope of the strategies to reposition bicycles can be enhanced to involve not only operational decisions, such as the routing, but also tactical and strategic decisions.

We reviewed the literature to identify papers related to strategies to reposition bicycles that not only involve the definition of collecting and delivering tours, but consider tactical or strategic decisions. We searched for articles in the databases

ScienceDirect and *Scopus* with the terms *"bicycle sharing system" AND (balancing OR repositioning)*. A total of 86 not repeated papers were retrieved. Based on a summary skimming, we focused on 14 of them that involve strategic and tactical decisions.

The design of the system is the most common topic that involves tactical and strategic decisions. Authors usually consider decisions such the number and location of the stations simultaneously with the inventory level of bicycles at each of them. Lin et al. (2013) study a strategic design problem formulated as a hub location inventory model and take into account not only the number and locations of bicycle stations, but also the inventory levels of sharing bicycles to be held at the stations. A similar problem is addressed by Nair and Miller-Hooks (2014) trough an equilibrium network design model that determines the optimal configuration of a vehicle sharing system, including bicycles. The problem is formulated as a bi-level mixed-integer program. At the upper level, the operator determines the optimal configuration of the system (supply). At the lower level, users respond to the system configuration and optimize their personal itineraries. Similarly, several author study strategic decisions at the system design level while considering operational aspects such as the service level (Çelebi et al., 2018), or the system demand characteristics (Martinez et al., 2012; Frade and Ribeiro, 2015). In Vogel (2016) a service network design approach is proposed to cover tactical planning decisions of BSS. It integrates mathematical optimization and intelligent data analysis to aggregate operational data. Relocation operations are anticipated by a dynamic transportation model that triggers relocation services between pair of stations. Neumann-Saavedra et al. (2015) extended this work to consider the service tours, that is, the sequence of the relocation services into tours.

The analysis of demand patterns at the operational level has been considered when tackling tactical and operational decision. Caggiani et al. (2018) proposes an optimization model to expand a BSS given a restricted budget with the objective of maximizing the global user satisfaction. They analyze historical usage patterns and use spatio-temporal clustering to address the need of adding or removing racks to each station, setting the optimal number of bikes, and deciding the need of building new stations. Zhang et al. (2017) determines the optimal inventory levels that need to be maintained at each bicycle station such that user dissatisfaction is minimized. The authors propose a new approach to estimate the user dissatisfaction and integrate it to bicycle repositioning and the vehicle routing, leading to a non-linear time-space network flow model. Faghih-Imani et al. (2017) analize not only the demand patterns but also the rebalancing operation patterns, to quantify and compare the influence of bicycle infrastructure attributes and land-use characteristics on: (i) demand, consisting of customer arrivals and departures, and (ii) rebalancing, consisting of the frequency and quantity of operator refills and removals. The results of two case studies (Barcelona and Seville) confirm that increasing the number of stations generates a reduction of the operator rebalancing needs. In addition, it showed that the presence of heterogeneous points of interest in each sub city district leads to lower requirements of rebalancing.

de Chardon et al. (2016) focuse on analyzing the rebalancing operational data and found that stations that are adjacent to transit hubs receive disproportionate amounts of rebalancing services in relation to the number of trips, and that rebalancing more often respond to morning and afternoon lack of parking docks rather than longer term accumulations of bicycles. Probabilistic and forecasting models have been used to estimate user patterns, taking into account various exogenous factors that influence the demand. Reynaud et al. (2018) developed a behaviorally quantitative model that allows system operators to forecast the potential problematic stations (full or empty). Understanding of the factors affecting bicycle availability will yield insights into the supply-and-demand mechanisms of Bicycle Sharing Systems, and allow the operators to better optimize their rebalancing procedures and/or plan the system modification (addition or relocation of stations and capacity).

The literature review supports our hypothesis that the analysis of operational data, trips between stations, and stock levels at the stations, can be used to guide tactical decisions such as the districting of the system network to facilitate the repositioning of bikes at the more congested hours of the operation.

4 Mathematical modelling

The districting problem falls within the strategic or tactical planning. It usually involves data from the day-to-day operation but focuses on the periods of higher demand (peak periods). The problem is solved to serve those periods and expected to be robust for the lower demand conditions (valley periods). This section describes the elements needed to face the districting problem while taking into account the demand patterns and stations importance.

First, we describe the methodology adapted from (Gaviria et al., 2016) to identify the peak periods based on the transactional information and to classify the stations depending on the criticality they have for the performance of the system. Then, we present a mathematical model for the districting problem in BSS that helps to understand the effect of considering demand patterns and stations importance when creating the repositioning zones.

4.1 Peak and valley periods and station priorization

Using data from EnCicla, the BSS of Medellín City, (Gaviria et al., 2016) proposed a methodology to identify the peak and valley periods and prioritize the stations. We use that methodology in this study using data from the system in Mexico City (ECOBICI). The methodology involves two elements, the definition of the peak hours for the operation of the system, as the districting problem would focus on that peak demand, and the prioritization of the stations regarding their influence in the performance of the system.

First, the operation shift is partitioned into small intervals. The length of the intervals is chosen using the percentiles of the distribution of the loans' duration, such that a large given percentage of the loans last less than the chosen value. The value is rounded, for the sake of practicality, to a meaningful value within the operational context (e.g., 15 or 30 minutes). The number of the transactions, loans and returns, in each of the intervals is calculated and statistical tests are run to validate the similarity of the demand patterns among the different days of the week. As suggested by authors such as (de Chardon et al., 2016), those patterns are usually similar for the week days and differ during the weekends. The distribution of the frequencies in the intervals is used to identify the peak and valley periods. Following what is observed in the literature at most three peaks are chosen (i.e., morning, noon and night), which corresponds to those that are above of a given percentile of the distribution of frequencies.

Once the peak intervals are identified, the information for those specific intervals is used to determine the priority of each station of the system. The priority is measured trough an index that involves the number of transactions and the unbalance of the station. The number of transaction of a given station T_i is calculated as the sum of bicycle loans and returns in the station i during the given interval, while the unbalance B_i of the station i is calculated as the difference between the number of bicycle loans and returns. The priority index p_i for a station i in the set of stations $\mathcal{E}$ is calculated as shown in equation (1). The value of the index is bigger for those stations that have a large unbalance between loans and returns and a high number of transactions.

$$p_i = \frac{T_i * B_i}{\sum_{j \in \mathcal{E}} T_j} \tag{1}$$

The percentiles of the priority index distribution are used to identify the level of priority of each station. The stations are classified in four different levels of importance based on where they are located in the distribution of the calculated priority index, as it is shown in Figure 3. For instance, the stations with the highest priority are those below the 5th percentile and above the 95th percentile of the priority index distribution.

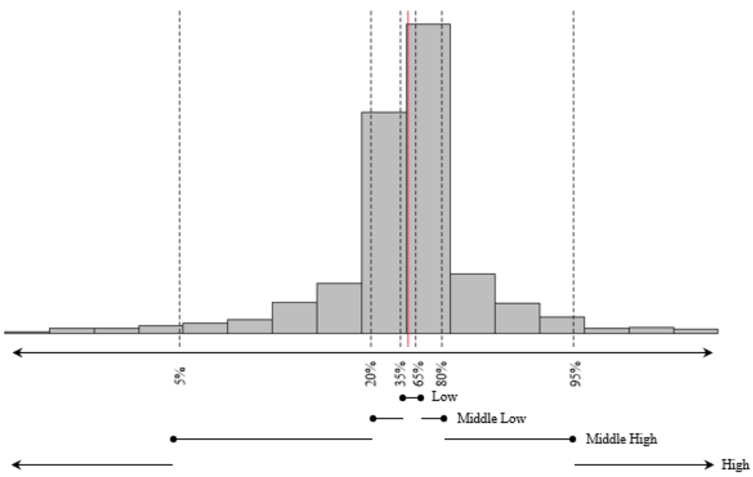


Fig. 3 Definition of the priority index levels of the stations.

4.2 Districting problem formulation

We propose a mathematical model based on the formulation for the capacitated clustering problem described in Koskosidis and Powell (1992). The model can be formally stated as follows:

Let $\mathcal{E}$ be the set of stations, $\mathcal{C}$ a subset of $\mathcal{E}$ that contains the candidate stations to be the center of a repositioning zone, and $\mathcal{P}$ a set of importance levels that defines the priority that each station is granted for the repositioning strategy. We define r_i^+ and r_i^- as the number of requests of bicycles and parking docks, respectively, that the station i has during the peak hour. The parameter d_{ij} is the distance between station i and j, and c_{ij} is a binary indicator of the connectivity between stations i and j. The indicator c_{ij} takes the value 1 if the distance between stations i and j is less than a given maximum distance, $dmax$, and the stations are accesible from each other, that is, there are not physical barriers between them. Additionally, p_{il} is a binary parameter that indicates whether the station i is assigned priority l, and k is the number of repositioning zones to be defined. The model considers two set of decision variables. The binary variable y_j indicates whether the candidate station j is designated to be the center of a repositioning zone, while the variable x_{ij} indicates whether the station i is assigned to the zone centered in the station j. The model is described as follow

$$\min \sum_{i\in\mathcal{E}} \sum_{j\in\mathcal{C}} d_{ij} x_{ij} \tag{2}$$

s.t.

$$\sum_{j\in\mathcal{C}} x_{ij} = 1 \qquad \forall i \in \mathcal{E} \tag{3}$$

$$x_{ij} \leq c_{ij} y_j \qquad \forall i \in \mathcal{E}\ \forall j \in \mathcal{C} \tag{4}$$

$$\sum_{j\in\mathcal{C}} y_j = k \tag{5}$$

$$\left| \frac{\sum_{i\in\mathcal{E}} r_i^+ x_{ij} - \sum_{i\in\mathcal{E}} r_i^- x_{ij}}{\sum_{i\in\mathcal{E}} r_i^+ x_{ij} + \sum_{i\in\mathcal{E}} r_i^- x_{ij}} \right| \leq \alpha \qquad \forall j \in \mathcal{C} \tag{6}$$

$$\left| \sum_{i\in\mathcal{E}} p_{il} x_{ij} - \left\lfloor \frac{\sum_{i\in\mathcal{E}} p_{il}}{k} \right\rfloor y_j \right| \leq \beta \qquad \forall l \in \mathcal{P}\ \forall j \in \mathcal{C} \tag{7}$$

$$x_{ij} \in \{0,1\} \qquad \forall i \in \mathcal{E}\ \forall j \in \mathcal{C} \tag{8}$$

$$y_j \in \{0,1\} \qquad \forall j \in \mathcal{C} \tag{9}$$

Objective function (2) minimizes the sum of the distances of each station to the center of the repositioning zone to which the station was assigned. It aims at generating compact zones. To that end, a function that minimizes the maximum distance between each pair of stations within the same repositioning zone would be preferred, because it explicitly involves the diameter of the repositioning zones (Kalcsics, 2015). However, linearizing that type of function for instances with hundreds of stations would affect the tractability of the model, as it considerably increases the number of constraints. Therefore, we opted for the function in (2) and included a maximum coverage distance in the definition of the connectivity c_{ij}. Constraints (3) to (5) establish the number of repositioning zones and ensure that each station is assigned to one of the created zones. Constraints (6) aim at balancing the bicycle and parking docks demand within each zone, such that at peak hours not all the stations require bicycles (or all them require to free parking docks). The parameter α is the maximum tolerable percentage of unbalance between the demand of bicycles and the demand of parking docks within any repositioning zone at the peak hour. Similarly, constraints (7) aim at distributing homogeneously the stations among the zones respect to the priority levels. That is, that the most critical stations are not concentrated in a reduced set of zones. The parameter β is the maximum difference allowed between the number of stations of a given priority level and the ideal value within each zone. The ideal value is the ratio between the number of stations of a given priority level and the number of repositioning zones. A different value of β might be used for each level of priority. Note that constraints (6) and (7) are not linear but can be easily linearized. Finally, constraints (8) y (9) define the variables to be binary.

5 Computational analysis

The model proposed in section 4.2 was used to understand how the creation of the repositioning zones is influenced by considering demand patterns and stations priority. This section describes the system that is used as a case study and the experiments and metrics to evaluate the different scenarios that were built. Then, results and insights obtained from the experiments are discussed.

5.1 Description of the case study

The system ECOBICI, that operates in Mexico City, was used as the case study to gather information and build the instances to run the computational experiments. ECOBICI was launched in 2010 and by the time of the study it had 452 stations distributed in an operational area of $35km^2$. The system has a policy of open data that allows the access to the operational data trough and API (https://www.ecobici.cdmx.gob.mx/en/). For this study, we consider the transactional record of loans between September and November of 2016 (2'353,389 travel records).

Figure 4 shows the hourly demand for the three months revealing that the demand patterns seem to be alike. A similar analysis shows that the demand patterns seem to be independent of the day of the week, except for the weekend days. Additionally, the peak hours and stations priority are defined following the methodology described in section 4.1. The operational time frame is divided in intervals of 30 minutes due to the fact that 92.0% of the trips last less than 32 minutes. Three peak periods were identified. However, we focused on the period of higher demand during the afternoon (i.e., around 18:00), as it accounts for around

30.0 % of the total demand of the system.

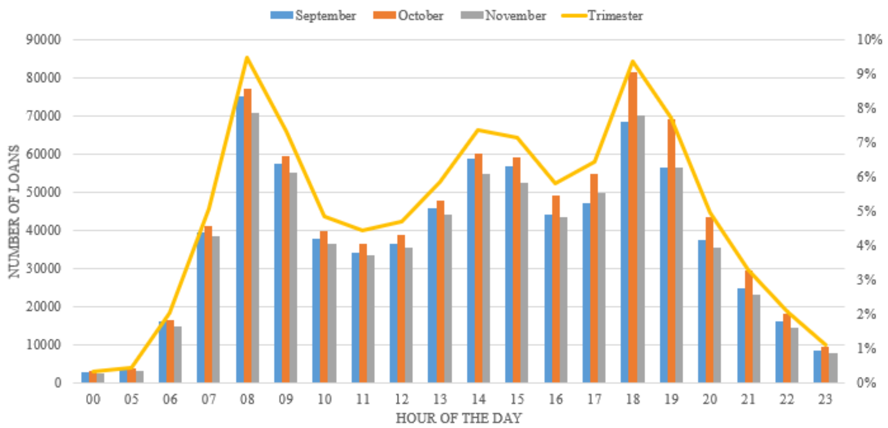


Fig. 4 ECOBICI hourly demand patterns.

Based on the data, three instances were created. A large instance that considers the 452 stations (*instance1: PT*), and two medium size instances that consider 224 stations (*instance2: PT-V1*) and 228 stations (*instance3: PT-V2*). Figure 5 shows the spatial distribution of those instances. For the *instance1*, the stations have to be partitioned into 15 repositioning zones, while *instance2* and *instance3* consider only 7 zones.

5.2 Design of experiments

The computational experiments were designed to address the question of *how considering demand patterns and stations criticality affect the definition of the repositioning zones of a BSS?*. Three analysis threads were defined: **i.)** Reveal the impact of demand patterns and stations priority on the compactness and spatial distribution of the repositioning zones **ii.)** Analize the sensitivity of the repositioning zones to different values of the parameters that enforce demand patterns and stations priority within the districting model. **iii.)** Evaluate the impact that considering the demand patterns and stations priority has in the difficulty to solve the associated mathematical model.

The three instances described in section 5.1 were used for the analysis. Several scenarios were built for each instance as a combination of the levels of the experimental factors. The factors considered were:

– The maximum distance, *dmax*, used to establish the connectivity c_{ij} of each pair of stations. This parameter helps on defining an upper bound for the diameter of the repositioning zones.

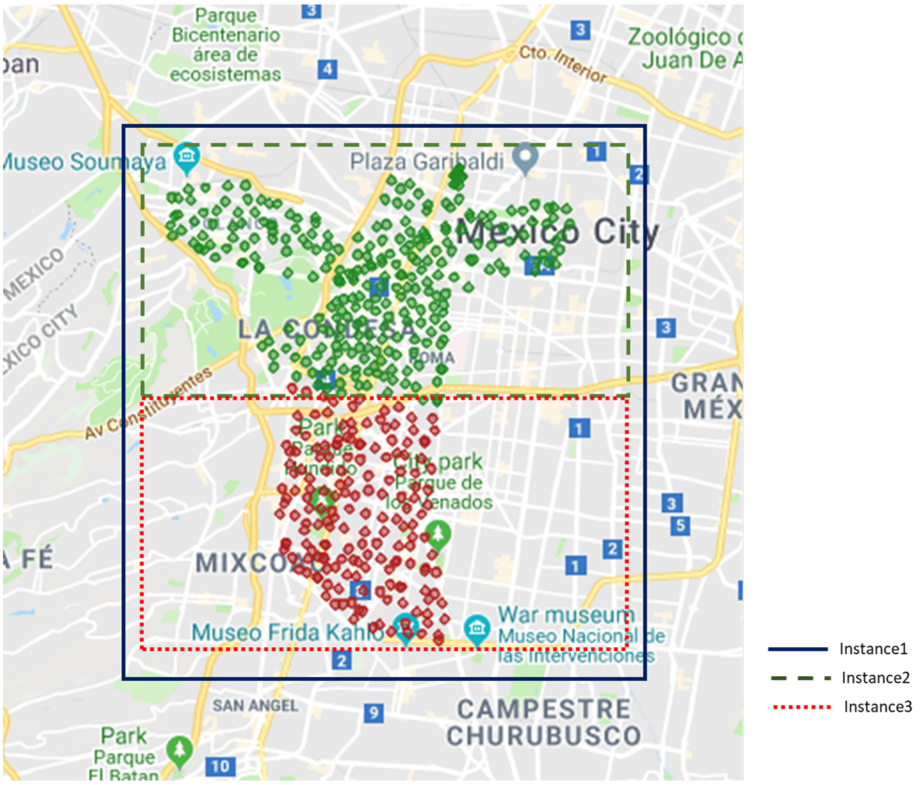


Fig. 5 Spatial distribution of the instances used in the computational experiments

- Parameter α, described in constraints (6) of the mathematical model, which represents the maximum tolerable percentage of unbalance between the demand of bicycles and the demand of parking docks within any repositioning zone at the peak hour.
- Parameter β, described in constraints (7) of the mathematical model, that represents the maximum difference allowed between the number of stations of a given priority level and the ideal value within each zone.
- A binary parameter, $allPrior$, for the optimization model, that establishes whether the constraints (7) that balance the distribution of the stations among the zones are applied to all priority levels or only to the stations with higher priority. That is, $allPrior = 1$ implies that a constraint to balance the number of stations of each priority level is considered in the model, while $allPrior = 0$ indicates that constraint (7) is added only for the stations with higher priority.

Table 1 summarizes the factors and their corresponding experimental levels. Considering the three instances and all possible combinations of these factors, a total of 432 experiments were run.

The results of each scenario were compared against the baseline model, that is, the model that does not consider the constraints to balance the demand of the zones (i.e., constraints (6)) and the distribution of the stations according their

Table 1 Experimental factors and levels

Factor	Level	Units
dmax	2500, 3500, 4500	meters
α	0.05, 0.1, 0.2, 0.3, 0.5, 0.7	deviation from the ideal (%)
β	2, 5, 10, 20	deviation from ideal (# of stations)
allPrior	0, 1	

priority level (i.e., constraints (7)). The results of each experiment, specifically the changes respect to the unrestricted model, were summarized through four indicators:

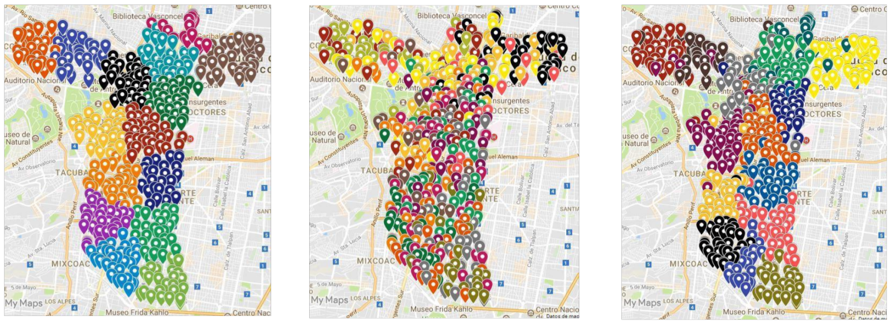
- $\%\Delta TotalDist$: Percentage change on the total sum of distances from the stations to the center of their zones.
- $\%\Delta AvgDiameter$: Percentage change on the average of the diameters of the repositioning zones.
- $\%\Delta MaxDiameter$: Percentage change on the maximum diameter of the repositioning zones.
- *cpuTime*: Computational time to solve the experiment.

5.3 Results and discussion

The model was implemented in Gurobi 7.5 using the python interface. It was run in a computer Intel Core i5 with 4GB of RAM. Each scenario was run either for 10.000 seconds or until reach a 2.0% of optimality gap. Not all combination of parameters are feasible. In total, 252 out of 432 scenarios were feasible, and 242 of them were solved to optimality within the defined time limit. All results were analyzed using the statistical software R.

We first analyze the impact of the demand patterns and stations priority on the compactness and spatial distribution of the repositioning zones. Figure 6 shows the results for the baseline case and two scenarios with different values of the parameters α and β. The results reveal that enforcing those constraints affects the compactness of the repositioning zones. Moreover, it is shown that the impact would depend on the value given to the design parameters. For some combination of parameters (e.g., $\alpha = 15$, $\beta = 5$) the cluster of the zone are not distinguishable, while other combinations of them generate well defined zones.

Having shown empirically that parameters α and β affect the design of the repositioning zones, we focus on understanding the effect that each of the factors has. Figure 7 shows that among the five factors considered, α and β have the largest impact on the objective function indicators. The other two factors associated to parameters of the model, namely *dmax* and *allPrior*, seem to have a reduced impact. In the case of the parameter *dmax*, it was used to define the connectivity c_{ij} between each pair of stations. The double of its value defines an upper bound for the diameter of any repositioning zone. The objective function incentivizes the zones to be as compact as possible but it does not consider the diameter of each



(a) **(b)** **(c)**

Fig. 6 Results for three scenarios: **(a)** baseline **(b)** $\alpha = 15$, $\beta = 5$ **(c)** $\alpha = 20$, $\beta = 8$

repositioning zone explicitly. It is observed that the impact on the objective function indicators related to the diameter of the zones (i.e., $\%\Delta AvgDiameter$ and $\%\Delta MaxDiameter$) is larger than to the one related to the sum of all distances. Regarding the parameter *allPrior*, the number of stations with the highest priority is usually less than the number of stations in other priority levels. Enforcing constraints (7) to the stations with highest priority (i.e., $allPrior = 0$) would be usually as restrictive as adding the constraints to each of the priority levels (i.e., $allPrior = 1$). That is the reason for this parameter to have a limited impact on the objective function indicators. Finally, the factor *instance* has a moderate effect on the indicators of the objective function, but it is not a parameter of the model and it is not controllable by the experimenter.

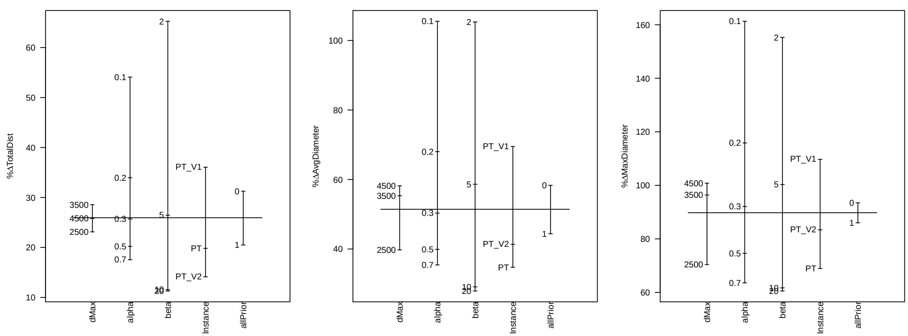


Fig. 7 Effect of the different factors over the objective function indicators

Table 2 presents the results of an ANOVA analysis to test the significance of the factors in the variability of the objective function. The basic model considered all the factors and their interactions, but the final model that is reported in Table 2 only considers those that are statistically significant. The analysis confirms that parameters α and β are also the responsible for most of the variability on the objective function indicators when the balancing constraints are including into the districting model. That is, when comparing to the case in which the constraints that involve those parameters are not taken into account, the objective function value and variability increase depending on the values given to α and β.

Table 2 ANOVA for the experimental factors

Factor	Df	Sum Sq.	Mean Sq.	F Value	pr(> F)
α	4	106.9	26.7	60.6	$< 2.2e^{16}$
β	3	114.1	38.0	86.3	$< 2.2e^{16}$
$allPrior$	1	3.3	3.3	7.5	0.007
$\alpha : \beta$	12	16.0	1.3	3.0	0.0006
$Residuals$	220	96.9	3.3	0.441	

Figures 8 and 9 show that as the value of α and β decrease, that means that constraints (6) and (7) are tighten, the objective function deteriorates, which is also reflected in the increment of the diameter of the repositioning zones. This effect seems to be stronger for values of α below 20%. Similarly, the objective function indicators increase drastically when the value of β is less or equal than 5, in that case an increment on the variability of the indicators is also observed. The effect of constraints (6) and (7) is more evident for the indicators $\%\Delta AvgDiameter$ and $\%\Delta MaxDiameter$ than for $\%\Delta TotalDist$. That is, the percentage of change in the average and maximum diameter when adding those constraints, respect to the baseline case in which the constraints are not included, is greater than for the sum of all distances in a cluster. The variability is smaller in the case of the $\%\Delta MaxDiameter$ because this indicator is less sensible to small changes in the distribution of the zones and it only varies when the zone with the maximum diameter changes.

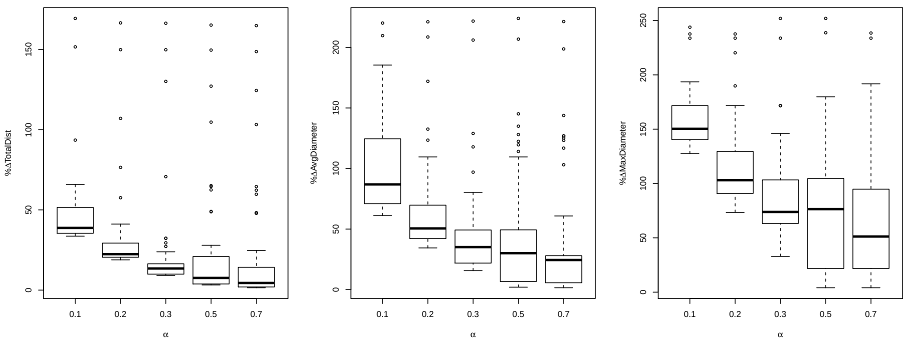


Fig. 8 Box plot of the objective function indicators vs the parameter α

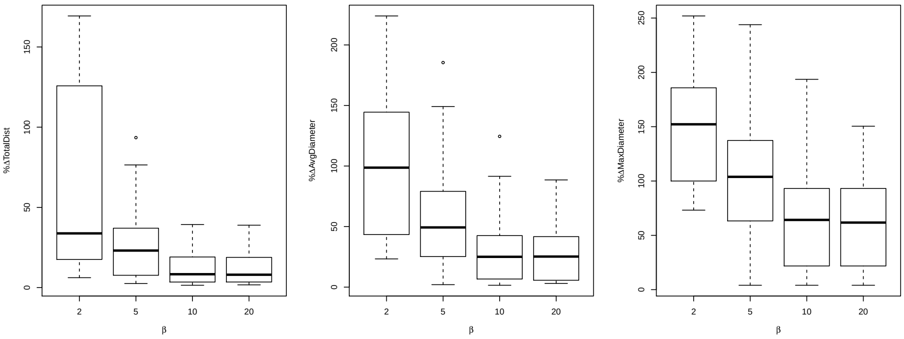


Fig. 9 Box plot of the objective function indicators vs the parameter β

Almost 60.0% of the unfeasible instances (i.e., 108 out of 180) were generated using either a value of $\alpha = 0.05$ or $\beta = 2$. That percentage increases to 90.0% when $\alpha \leq 0.1$ or $\beta \leq 5$. These results show that when enforcing the balance through constraints (6) and (7), both constraints can not be set to the minimum value at the same time. Additionally, the analysis of interaction between the factors presented in Figure 10 shows that the combined effect of the parameters α and β is intensified when those parameters are tighten simultaneously. Therefore, there is trade-off between the parameters that has to be taken into account when involving the balancing constraints into the model. The figure also shows that constraints (7) do not have a significant effect on the indicators of the objective function when β is set to values greater than 10. However, below that value, particularly when β is set to 5, the increment on those indicators is considerable. Based on the analysis, we suggest to set parameter α to a value around 2.00% while β should be fixed to a value between 5 and 10.

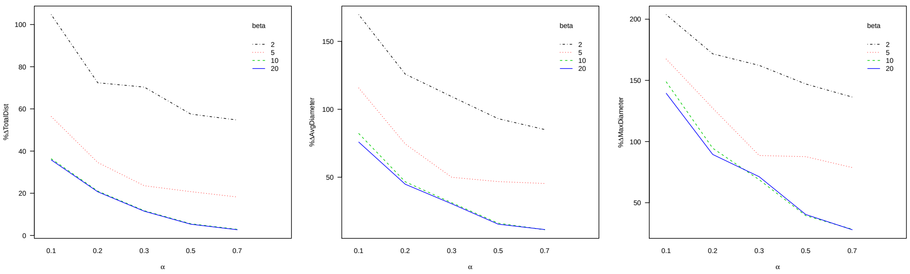


Fig. 10 α vs β interaction plots for the different objective function indicators

We now focus on analyzing the impact that considering the demand patterns and stations priority has in the difficulty to solve the associated mathematical model, specifically on the computational time to solve it. Figure 11 shows that the parameter β has the largest effect on increasing the CPU time followed by the parameter α, as those parameters are tighten the computational time increases. The factor *instance* also has a significant impact, as the largest instance (PT) demands more computational time. Although the parameter *allPrior* does not seem to affect the objective function, it seems to have an effect on the computational time. As expected, the model that considers the set of constraints (7) for all the priority levels is more demanding of computational resources due to the large amount of constraints that it involves. Table 3 presents the average and maximum computational time for each of the combinations of the parameters α and β. There is not values for $\alpha = 0.05$ as most of the instances were unfeasible or not solved within the time limit. Similarly, only few instances were feasible when $\alpha = 0.1$, therefore the computational time of those instances are not comparable with the other combinations of the parameters. In general, the computational time increases significantly when the parameter β is tighten from 10 to 5, while the computational time increases moderately as the parameter α decreases.

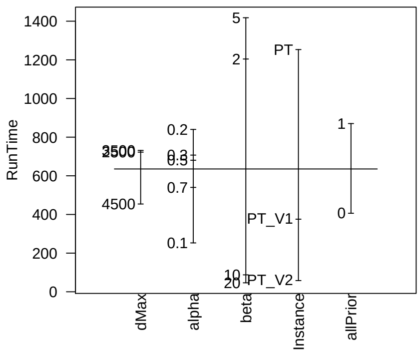


Fig. 11 Effect of the different factors on the computational time

Table 3 Computational time for the different combinations of parameters α and β

α	β = 2 Avg.	β = 2 Max.	β = 5 Avg.	β = 5 Max.	β = 10 Avg.	β = 10 Max.	β = 20 Avg.	β = 20 Max.
0.1	184.0	399.5	768.0	4331.6	19.2	29.0	17.5	28.0
0.2	1123.1	6660.3	1988.2	8684.2	321.6	2596.3	117.0	525.2
0.3	1420.7	9064.9	1719.7	5002.8	95.5	274.9	83.7	253.7
0.5	1390.2	10000.2	1451.9	8190.9	21.1	86.9	17.2	43.0
0.7	1232.7	8105.6	1082.5	5719.3	15.6	38.7	13.4	35.8

6 Conclusions

This paper addresses the problem faced by BSS operators when they have to create zones to be served by the different repositioning vehicles. This is a tactical decision that affects some of the most studied operational problems (e.g., routing and inventory management). Therefore, studying the districting problems for Bicycle Sharing Systems provides a broader perspective for the problematic of how to balance their operation, and could facilitate the implementation of repositioning strategies. We argue, based on the literature and the interviews with operators, that the districting problem should not be based exclusively on proximity and accessibility but should take into account additional criteria such as demand patterns and stations hierarchy.

We propose a mathematical model that involves constraints to balance the bicycle and parking docks demand within each zone during a peak hour. It also distributes the stations among the zones homogeneously respect to the priority levels. The model was tested using instances built with real operational data from the system ECOBICI in Mexico City. The results confirm that including those constraints has an impact on the compactness of the repositioning zones. The sum of the distances from the stations to the center of their zones increases as the constraints are tighten, so does the diameter of the zones. There is a trade-off between the parameters α and β, because the interaction between small values of them generates a significant increment on the objective function and the computational time. Based on the analysis, we suggest to set parameter α to a value around 20% while β should be fixed to a value between 5 and 10. Smaller values for those parameters usually leads to infeasible instances which means that the ideal of perfect balance, enforced by constraints (6) and (7), can not be met.

The proposed model solved to optimality 96.0% of the feasible instances within a time limit of 10.000 seconds. The largest of those instances consider 452 stations which is representative for a medium size BSS (Citi Bike, the system of New York, counts with around 750 stations by the end of 2018). The districting problem models a tactical decision which is usually taken once or twice a year. Therefore, the results allows us to say that computing times are acceptable for this kind of decision making process.

Finally, as pointed out when we described the model in section 4.2, it is worth to try alternative objective functions, as those described in (Kalcsics, 2015). For instance, minimizing the maximum distance between each pair of stations within the same repositioning zone. Although, some of those functions would account better for the compactness of the repositioning zones, in many cases they are non-linear and would demand a large number of constraints or modifications on the formulation that would affect the performance of the model. In that case, the use of approximate approaches, such as metaheuristic methods, is a promising research direction for future work.

References

Arroyave, M., 2016. Design of strategies for the repositioning of units in Bicycle Sharing Systems. Master's thesis, Universidad de Antioquia.

Büttner, J., Petersen, T., 2011. Optimising bike sharing in european cities-a handbook.

Caggiani, L., Camporeale, R., Marinelli, M., Ottomanelli, M., 2018. User satisfaction based model for resource allocation in bike-sharing systems. *Transport Policy*

Çelebi, D., Yörüsün, A., Işık, H., 2018. Bicycle sharing system design with capacity allocations. *Transportation Research Part B: Methodological* 114, 86–98.

de Chardon, C.M., Caruso, G., Thomas, I., 2016. Bike-share rebalancing strategies, patterns, and purpose. *Journal of transport geography* 55, 22–39.

Dell'Amico, M., Hadjicostantinou, E., Iori, M., Novellani, S., 2014. The bike sharing rebalancing problem: Mathematical formulations and benchmark instances. *Omega* 45, 7–19.

Faghih-Imani, A., Eluru, N., El-Geneidy, A.M., Rabbat, M., Haq, U., 2014. How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal. *Journal of Transport Geography* 41, 306–314.

Faghih-Imani, A., Hampshire, R., Marla, L., Eluru, N., 2017. An empirical analysis of bike sharing usage and rebalancing: Evidence from Barcelona and Seville. *Transportation Research Part A: Policy and Practice* 97, 177–191.

Frade, I., Ribeiro, A., 2015. Bike-sharing stations: A maximal covering location approach. *Transportation Research Part A: Policy and Practice* 82, 216–227.

Gaviria, B., Gómez, M., Rodríguez, M., 2016. Demand caracterization of EnCicla Bike Sharing System. Master's thesis, Universidad de Antioquia.

Kalcsics, J., 2015. Districting problems. In *Location science*. Springer, pp. 595–622.

Koskosidis, Y.A., Powell, W.B., 1992. Clustering algorithms for consolidation of customer orders into vehicle shipments. *Transportation Research Part B: Methodological* 26, 5, 365–379.

Laporte, G., Meunier, F., Calvo, R.W., 2015. Shared mobility systems. *4or* 13, 4, 341–360.

Lin, J.R., Yang, T.H., Chang, Y.C., 2013. A hub location inventory model for bicycle sharing system design: Formulation and solution. *Computers & Industrial Engineering* 65, 1, 77–86.

Martinez, L.M., Caetano, L., Eiró, T., Cruz, F., 2012. An optimisation algorithm to establish the location of stations of a mixed fleet biking system: an application to the city of Lisbon. *Procedia-Social and Behavioral Sciences* 54, 513–524.

Miranda, P.A., González-Ramírez, R.G., Smith, N.R., 2011. Districting and customer clustering within supply chain planning: a review of modeling and solution approaches. In *Supply Chain Management-New Perspectives*. IntechOpen.

Nair, R., Miller-Hooks, E., 2014. Equilibrium network design of shared-vehicle systems. *European Journal of Operational Research* 235, 1, 47–61.

Neumann-Saavedra, B.A., Vogel, P., Mattfeld, D.C., 2015. Anticipatory service network design of bike sharing systems. *Transportation Research Procedia* 10, 355–363.

Reynaud, F., Faghih-Imani, A., Eluru, N., 2018. Modelling bicycle availability in bicycle sharing systems: A case study from Montreal. *Sustainable cities and society* 43, 32–40.

Salazar-Aguilar, M.A., Ríos-Mercado, R.Z., Cabrera-Ríos, M., 2011. New models for commercial territory design. *Networks and Spatial Economics* 11, 3, 487–507.

Vogel, P., 2016. Service network design of bike sharing systems. In *Service Network Design of Bike Sharing Systems*. Springer, pp. 113–135.

Zhang, D., Yu, C., Desai, J., Lau, H., Srivathsan, S., 2017. A time-space network flow approach to dynamic repositioning in bicycle sharing systems. *Transportation research part B: methodological* 103, 188–207.